Biotechnology Divsion - Table of Contents

Biotechnology Division (831)

G.L. Gilliland, Chief

A. Division Overview

The Biotechnology Division is the focus of the NIST effort addressing critical measurement and data needs for the rapidly developing biotechnology industry. The mission of the Biotechnology Division is to provide measurement infrastructure necessary to advance the commercialization of biotechnology. This is achieved by developing a scientific and engineering technical base along with reliable measurement techniques and data to enable U.S. industry to produce biochemical products quickly and economically with appropriate quality control. The Division has established a variety of long-range research programs to maintain critical expertise needed for the development of advanced measurement methods, Standard Reference Materials and databases for use by industry and other research enterprises. It fosters collaboration among NIST scientists conducting biotechnology research, and raises the visibility of the NIST Biotechnology program, which leads to enhanced collaborations with industry, universities and other government agencies.

To plan and guide Division research programs effectively, Division scientists are active in many forums that provide feedback on the long-term and short-term scientific and technological needs for biotechnology commercialization. These activities supplement the high level of Division participation in scientific meetings and topical workshops. For example, during FY01 the Division continued active participation in the Biotechnology Industry Organization (BIO), in the IUPAC Commission on Biophysical Chemistry, and in the ASTM Committee E-48 on Biotechnology. Division members were also active participants in a number of important workshops held at NIST. The Division also worked closely with the NIST Advanced Technology Program (ATP) in a variety of ways. Division members served as NIST technical representatives in the annual review of a number of active ATP grants. They also presented the results of Division research programs in ATP-sponsored workshops and public meetings. Additionally, Division members have been involved in fundamental research that is directly related to ATP focus areas, and also involved in research directly related to specific ATP projects. Throughout the year, the Biotechnology Division continued to play a major role in the planning and coordination of Federal biotechnology research through memberships in the working groups associated with the Subcommittee on Biotechnology of the National Science and Technology Council (NSTC). Involvement in these activities assures that the Biotechnology Division is an active participant in the decision making process for prioritizing and directing funding of federal research especially in areas that impact the private sector.

The staff of the Biotechnology Division consists of 49 NIST employees and a comparable number of contract researchers, guest scientists, and postdoctoral fellows. The Division is organized into four groups: (1**) DNA Technologies**; (2) **Bioprocess Engineering**; (3) **Structural Biology**; and (4) **Biomolecular Materials**. In addition the division is continuing to develop a program in **Bioinformatics** independent from the other group efforts. A brief overview of the activities of each Group in the Division, and highlights of several research programs, are given below:

The DNA Technologies Group has focused on research efforts to meet goals and objectives in areas of data dissemination, measurement science, and standards. In the area of DNA diagnostics, the group has established the NIST-Early Detection Research Network Biomarker Validation Project, an integral part of an NCI large-scale effort involving, academic, clinical, industry, and government laboratories to develop new biomarkers for early cancer detection. As part of this effort a NIST-NCI Workshop was held at NIST on August 30 and 31, 2001, that focused on *early cancer detection and nanotechnology*. The workshop focused on new technology and interdisciplinary research that will play an important role in the discovery, validation and clinical implementation of novel cancer biomarkers. Current validation research is focused on the detection of chromosome breakage, telomerase activity in serum, and the detection of homoplasmic mitochondrial changes in cancer. This final activity is complementary to other ongoing programs addressing the mitochondrial genome, as described below. Other efforts in the area of DNA have the goal of providing the clinical diagnostics community with accurate protocols and measurements for the detection of genetic disorders. For example, Fragile X syndrome is the leading heritable cause of mental retardation; thus, accurate quantification of the number of repeat sequences is necessary for carrier screening in family planning as well as for accurate diagnosis of disease and pre-natal screening. Standardization of PCR amplification for Fragile X Trinucleotide Repeat measurements is an ongoing effort in the DNA Technologies Group. Further research efforts focus on robust amplification of full mutation alleles, and the stability of triplet repeats of all sizes in plasmid vectors. This research is expected to result in optimized laboratory protocols to provide accurate measurement capabilities for triplet repeats. The primary customers are the clinical diagnostics community and those genetic testing laboratories that screen the DNA of patients for triplet repeat diseases; specifically, fragile X syndrome.

The DNA Technologies Group's efforts in measurement science for emerging fields of diagnostics and therapeutics include the study of the mechanisms of oxidative stress, DNA damage and actions of DNA repair enzymes. These efforts include the development of sensitive measurement techniques for this purpose and for the measurement of oxidative stress biomarkers. Oxidative stress is the harmful condition that occurs in living cells when there is an excess of free radicals, a decrease in antioxidant levels, or both, and is associated with many diseases, such as cancer, heart disease and Alzheimer's disease. Using mass spectrometric techniques that include gas chromatography/mass spectrometry and liquid chromatography/mass spectrometry, the group has an ongoing program on the characterization of DNA damage and the functions of numerous DNA repair enzymes in living cells. The program also includes studies on the mechanism of action of redox-cycling compounds that have potential use in chemotherapy. To ensure the accuracy, traceability, and comparability of measurement results among laboratories obtained by mass spectrometric techniques, a standard reference material (SRM) consisting of twelve stable isotope-labeled modified DNA bases is being developed for the diagnostic and research community. As part of these efforts, a NIST co-sponsored international conference was held on April 2-5, 2001. This conference entitled "**Second International Conference on Oxidative Stress and Aging: Technologies for Assessment and Intervention Strategies**" was attended by over 200 people throughout the world and focused on the latest research on molecular gerontology with emphasis on oxidative stress-related diseases and the need and criteria for the development of measurement standards, guidelines and best practices in measuring parameters of oxidative stress in humans.

Another area of research is the identification of cellular molecular biomarkers that can be used for quality assurance and quality control of tissue-engineered materials in terms of genetic

damage. Tissue engineering often involves the mixing of cells with various polymers that may cause genetic damage to the cell. In order to assure that tissue-engineered materials are free of genetic changes that could occur during the *in vitro* development phase, cellular biomarkers are being identified that could be used as safety biological checkpoints by industry during the manufacturing of these products. Areas of investigation include oxidative stress monitoring in terms of oxidative DNA damage, mitochondrial damage detection, genomic point mutation screening, and chromosome loss. In the same context, a specially modified capillary electrophoresis instrument is being used to study mechanisms of single-strand conformation polymorphism (CE-SSCP) and to provide optimized methods and standards that permit automated mutation detection.

Other efforts focus on the design and use of a peptide nucleic acid for detecting the heteroplasmic low-frequency mitochondrial DNA mutation associated with the disease MELAS (Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-like Episodes). A multitude of mitochondrial DNA (mtDNA) diseases have been correlated with single nucleotide polymorphisms (SNPs), mutations, insertions, and deletions. Most of these diseases are neuromuscular, but other diseases such as diabetes and cardiac insufficiency, also occur. Most pathogenic human mtDNA mutations are heteroplasmic meaning that the mutant mtDNA coexists with the normal mtDNA in the cell. A methodology is used to detect low levels of the single base pair heteroplasmic MELAS mutation. A series of peptide nucleic acids (PNAs) was designed to bind to the wild type mtDNA, thus reducing the polymerase chain reaction (PCR) amplification of the wild type mtDNA to background levels while permitting the mutant DNA to become the dominant product and readily discernable. This methodology permits detection of the MELAS mutation in asymptomatic or symptomatic carriers with low to undetectable blood levels of this mutation. Furthermore, matrix-assisted laser desorption/time-of-flight mass spectrometry (MALDI-TOF MS) is being used to directly examine and compare the proteomes of different bacterial species with the objective of trying to produce cellular spectral 'fingerprints' to identify the various bacteria. The effects of different cell growth conditions and cell growth stage on the spectral fingerprints of the whole cell samples are also examined. It has been demonstrated that freshly grown gram-positive bacteria, and gram-negative bacteria have unique MS spectra and that storage of samples at $-20^{o}C$ for two weeks does not affect the spectra. Although the stage of growth (logarithmic, stationary, decay) does significantly change the spectral fingerprint, it appears that there are characteristic identifying signals.

In the area of data dissemination, the NIST Web-based forensic database for Short Tandem Repeats (http://www.cstl.nist.gov/biotech/strbase) has come to be a popular web site. Over 44,000 hits have been recorded since implementation in October 1997. Recent upgrades to the site have added new information on rare variant alleles. This has become an essential touchstone for crime laboratories that need to know if their results are meaningful, or if they are from possibly spurious measurements. The database now includes over fourteen hundred STR references. The group has continued to provide analysis of forensic data for the College of American Pathologists and the National Institute of Justice. Programs in blind proficiency testing for the FBI and a quality assurance program for the Armed Forces DNA Identification Laboratory round out efforts that provide evaluated data and information.

Efforts also focus the development and evaluation of high throughput technologies for typing single nucleotide polymorphisms (SNPs) for human identification purposes. SNPs are the most common form of genetic variation in the human genome. The typing of SNPs throughout the genome can facilitate genetic mapping, disease association studies, and evolutionary studies. Recent analysis of SNPs markers located on the non-combining region of the Y chromosome provides information on tracing human migration patterns and evolution. Primer extension

assays to type SNPs located on the Y chromosome as well as in the mitochondrial genome are being designed in order to evaluate their usefulness in forensic applications. The results of these primer extension reactions are being analyzed using MALDI-TOF MS due to its inherent speed and accuracy for typing SNPs. The speed and accuracy of MALDI-TOF MS also allows rapid development of large DNA typing databases and population studies. This resulted in tools for the rapid optimization of multiplexed PCR and primer extension reactions to improve throughput for SNP analysis.

In an effort to support the growing demand for Y chromosome testing, an SRM is being developed that can be used to calibrate instrumentation and verify assay performance with Y STR and Y SNP markers. Over 140 candidate DNA materials for SRM 2395 have been screened. From these samples, five male and one female DNA samples have been selected and extensively characterized. These candidate materials reflect multiple alleles at each Y STR and SNP locus. These six candidate SRM samples with commercially available and research multiplex assays are being tested. In addition, all of the DNA samples are being sequenced to confirm exact repeat compositions of the STR markers. SNP markers are being analyzed with multiple technologies to confirm the polymorphic nucleotide present in each sample.

The **Bioprocess Engineering Group** is focused on the development of measurement methods, databases, and generic technologies related to the use of biomolecules and biomaterials in manufacturing. Measurement methods and data have been developed in the protein **biospectroscopy** area that will lead to improved understanding of intra- and inter-protein electron transfer processes. This understanding helps industrial biocatalyst development through more efficient utilization of carbon sources (e.g., renewable resources) and nutrients, and in developing new ways to drive organic syntheses such as the stereospecific hydroxylation of pharmaceutical precursors. The group's efforts to develop fluorescence intensity standards have resulted in the first standard reference material (SRM 1932) for measuring fluorescence intensity and new efforts to develop a particle fluorescence intensity standard (SRM 1933) to be used in application areas that involve cell sorting using flow cytometers and genomic microarray technology. These efforts are in response to industry and government needs as reflected by the output from recent workshops sponsored and cosponsored by the Biotechnology Division.

In the **biothermodynamics** of enzyme-catalyzed reactions research area, chromatography and microcalorimetry measurements have been combined with chemical equilibrium analysis to develop thermodynamic data for several industrially important biotransformations. The metabolic pathway by which microorganisms and plants convert glucose to aromatic amino acids is a current focus of the biothermodynamic measurements. This chorismate metabolic pathway is under current investigation by several large chemical companies as an environmentally friendly source of aromatic hydrocarbons. In the **preparative bioseparations** project, electrochromatographic/electrophoretic separation equipment and methodology are being applied to the separation of different physical forms of DNA (supercoiled plasmid, relaxed circular plasmid, linear genomic). Large-scale purification of these materials is an emerging need in industries developing gene therapies and diagnostic materials. This project is being phased out to make resources available for the emerging efforts to develop reference materials and methods for use in detecting **genetically modified grains**. This new effort is a collaboration among investigators at NIST and the Grain Inspection, Packers and Stockyards Administration (GIPSA) of the U.S. Department of Agriculture, and the Institute for Reference Materials and Measurements (IRMM) of the European Commission Joint Research Center.

Research projects in **biocatalytic systems** focus on enzyme characterization by site-directed mutagenesis, X-Ray diffraction of protein crystals and computational chemistry. These

techniques are being used to address focused, industrially important biotransformation problems such as those found in hydroxylation and aromatic amino acid metabolic pathways.

Research by the **Structural Biology Group** at the **Center for Advanced Research in Biotechnology** (CARB) is focused in four key areas of industrial biotechnology. These include: macromolecular structure determination by X-ray crystallography; molecular structure and dynamics elucidation by modern, high-field nuclear magnetic resonance (NMR) spectroscopy; physical, molecular and cellular biochemistry; and computational biotechnology and modeling. A balanced program in these four areas has been developed at CARB by recruiting a highly interactive group of scientists through both the University of Maryland Biotechnology Institute (UMBI) and NIST with interests and expertise in the theory and measurement of macromolecular structure-function relationships that underlie virtually all biological processes. CARB has established profitable interactions with several outside organizations in response to its distinct responsibility as part of a unique collaboration among NIST, the University System of Maryland and Montgomery County, MD. The CARB staff has also engaged in several technical activities that relate directly to NIST's core mission in measurement science and chemical and process information.

The research in the Structural Biology Group encompasses a broad range of activities. In the area of macromolecular structure determination, there were three principal areas of focus during the last year. Elucidation of the structure and function of enzymes in the chorismate metabolic pathway continues to be a key interest of the X-ray crystallography and modeling groups. This effort, part of the Division-wide activity aimed at the vitally important industrial area of **metabolic engineering**, aims to account for the high specificity and unusual activities of key biosynthetic enzymes in terms of unique structural attributes and theory. The work this year has yielded a better understanding of the mechanisms of enzymes in the chorismate biosynthetic pathway. An interesting new structure for aminodeoxychorismate synthase has revealed unique clues on the mechanism of *p*-aminobenzoate biosynthesis, and also shed some unexpected light on the evolution of chorismate utilizing enzymes that will allow for more rational approaches to engineer the structures and activities of this industrially important family of enzymes. Additional structural studies on mutational variants of aminodeoxychorismate lyase have provided important new data for computational studies of the reaction mechanism. The synergy between the structural, computational, enzymology groups has led to a new model for the mechanism of this interesting enzyme.

The Structural Biology Group is responding to the critical, new area of **structural genomics** by developing high-throughput approaches for elucidating the structures and functions of all the proteins encoded by entire genomes. The focus of this effort is to determine the structures for 'hypothetical' proteins of microbial genomes that may be useful drug targets for industry. This effort is not merely a 'scaling up' exercise, but is requiring the development of entirely new approaches to meet the formidable challenges involved in the rapid determination of a large number of macromolecular structures. Interestingly, the structures of several proteins have revealed new folds, and analyses of these structures have suggested new biological functions that are currently being explored using traditional biochemical assays.

In addition to its effort in high-resolution structure determination, the NMR group has expanded its activities by evaluating new methods for screening small molecule ligand binding to biochemical targets. This endeavor is related to industrial challenges to identify new drugs using **high-throughput screening**. The main area of interest is in small molecule ligands that disrupt physiologically important interactions between proteins and RNA. Because of their fundamental

role in the pathology of many infectious agents, as well as because of their unusual three-dimensional structures, these nucleic acid-protein complexes have become an important new target for clinical intervention. New methods to assess the effect of compounds on the maturation of the HIV virus may provide a profitable approach for identifying promising new therapeutics.

The activities in the biochemical group are primarily focused on analyzing with important macromolecular interactions using structural, mechanistic, spectroscopic and thermodynamic information. These activities aim to put descriptive, qualitative biochemical assays on a more quantitative, chemical framework. Current work on the structures and interactions of key recognition elements in several members of G-coupled protein receptors have suggested new, quantitative models for signal transduction pathways in vision and viral infection. The new approaches that have been developed to probe the interactions of these receptors with their ligands have suggested possible new routes for the **high-throughput screening** of compounds that inhibit their association, and thereby block cellular activation.

A new activity in computational biology that is having an impact in **bioinformatics** involves the incorporation of mutational biases in multiple protein sequence alignments. Since one of the first clues in identifying the function of a newly identified protein from genome sequencing is to compare its sequence against established databases, new tools that enable more accurate comparisons, or that can detect significance between distantly related protein families, should have a major impact on the annotation of the genome sequences of organisms.

The **Biomolecular Materials Group** is examining the structure and function of biological molecules at interfaces. This is done using chemically controlled surfaces that are engineered for specific biomolecular interactions. These systems are critical components of biosensors, bioelectronics, biocatalytic systems, and many diagnostic devices. **Biomolecular materials** thus enable diverse applications including pharmaceutical development, health care, environmental pollution monitoring, and chemical manufacturing. One focus of the group is the characterization of rugged artificial membrane system that mimic cell membranes. This **supported bilayer membrane** is composed of both artificial and natural lipid components. Development of the chemistries needed to attach the lipids to a surface has been essential to the progress in this area.

Fundamental understanding of the structural characteristics of biological molecules such as cell membrane receptors, optically active proteins and redox enzymes in this and other biomimetic matrices, and the relationship between structure and functional activity, are key to successful commercial applications. Molecular details are provided by **vibrational spectroscopies** such as infrared and surface plasmon resonance enhanced Raman. Models based on electromagnetic wave theory allow the calculation of simulated spectra that aid data interpretation. Recent development of **infrared ellipsometry** has been useful for determining lipid orientation and conformation. Similarly, **neutron reflectivity** (in collaboration with the NIST Neutron Research Facility) is providing high-resolution data of these membranes and the location of protein complexes in them. A**tomic force microscopy**, is being developed to investigate the molecular details of synthetic membranes.

Pore-forming proteins, such as **α-hemolysin,** are used as model systems to investigate how membrane protein structural conformational changes lead to functional changes. **Sensor and nanotechnology** application developments demonstrate the potential commercial usefulness of these proteins and protein mimics. Synthetic chemistry is providing novel compounds that improve assembly of matrices that may lead to nanotechnology standard reference materials.

In FY01 **Bioinformatics** program efforts concentrated on data uniformity and development of the physical archive of the NSF/DOE/NIH supported **Protein Data Bank** NIST partners in the **Research Collaboratory for Structural Bioinformatics** (RCSB) include groups from Rutgers University, the University of California San Diego Supercomputer Center. The central focus of the efforts within the Division is on improving the querying capability of the database resource by improving data uniformity. First stage of data uniformity work on approximately 15000 legacy Protein Data Bank (PDB) entries is completed and the results are now distributed to users (http://pdb.nist.gov). These data uniformity efforts focused on a broad range of issues like data quality, data standards and data exchange.

Up until now most of the efforts in uniformity and standards in bioinformatics and structural biology have focused on the data. Recent developments in technology have created an urgent need for similar efforts in software standards as well. The **Bioinformatics Software Resource** (BISR) is being designed to address this need and plans for a modern database for the purpose of public archival and distribution of software and tools in **Bioinformatics** are being implemented. Through this effort, the Biotechnology Division goal is to catalyze and contribute to the development of standard application program interface (API) for format conversions and data exchange between software. A prototype for a web resource for software in Bioinformatics has been developed (http://bioinfo.nist.gov/BISR/). The objective of this resource is to provide a stable archival and distribution facility for publicly available software in Bioinformatics. As a first step, our web resource provides key-word searchable links to web resources that distribute software.

Genotyping Single Nucleotide Polymorphisms in the Y Chromosome and the Mitochondrial Genome

CSTL Program: DNA Technologies
Authors: *P.M. Vallone, and J.M. Butler*

Abstract: Single nucleotide polymorphisms (SNPs) are the most common form of genetic variation in the human genome. SNPs exist in approximately 1 out every 1000 base pairs. The typing of SNPs throughout the genome can facilitate genetic mapping, disease association studies, and evolutionary studies. Recent analysis of SNPs markers located on the non-combining region of the Y chromosome provides information on tracing human migration patterns and evolution.

We are designing primer extension assays to type SNPs located on the Y chromosome as well as in the mitochondrial genome in order to evaluate their usefulness in forensic applications. The results of these primer extension reactions are being analyzed using matrix assisted laser desorption-ionization time of flight mass spectrometry (MALDI-TOF MS) due to its inherent speed and accuracy for typing SNPs. The speed and accuracy of MALDI-TOF MS also allows rapid development of large DNA typing databases and population studies.

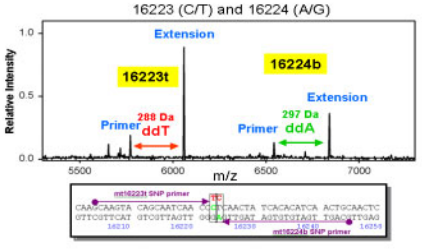
Our work has resulted in tools for the rapid optimization of multiplexed PCR and primer extension reactions to improve throughput for SNP analysis. Further, to date, we have compared various primer extension assays amenable to mass spectrometric analysis for SNP genotyping. The utility of MALDI-TOF MS to accurately and rapidly type samples is illustrated through results of Y chromosome and mtDNA SNP markers, including M9, M42, M45, M89, and M96.

Purpose: Development and evaluation of high throughput technologies for typing SNPs for human identification purposes.

Major Accomplishments: Collaborations with Thomas Parsons (AFDIL) and Michael Hammer (U of AZ) in order to test novel markers located in mitochondrial genome and Y chromosome. Poster of SNP typing technology presented at the 12th International Symposium on Human Identification Biloxi, MS October 2001.

Impact: Our efforts in evaluating technologies assists other labs and agencies that are currently involved in typing SNPs.

Future Plans: Developing assays that will also be compatible with other instrumentation formats such as capillary electrophoresis, fluorescence polarization, and fluorescent microspheres.



Results of genotyping (by MALDI-TOF MS) two adjacent SNPs in the control region of the mitochondrial genome (16623 and 16224). The mass difference between the primer and extension products determines the genotype. Two unique primers (depicted below) were designed to bind the 'top' and 'bottom' strands of the template in order to successfully type both polymorphisms in the same reaction.

Development of Y STR Megaplex Assays

CSTL Program: DNA Technologies
Authors: *J.M. Butler, P.M. Vallone, M.C. Kline, and R. Schoske*

Abstract: Y chromosome short tandem repeat markers have a number of applications in human identity testing including typing the perpetrator of sexual assault cases without differential extraction and tracing paternal lineages for missing persons investigations. In order for Y STR systems to become more widely accepted within the forensic DNA typing community, robust multiplex assays are required. We have focused on the design and development of new Y STR multiplexes.

We are working with 22 Y STR markers including: the hexanucleotide repeat DYS448; pentanucleotides DYS438, DYS446, DYS447, DYS450; tetranucleotides DYS19, DYS385I/II, DYS389I/II, DYS390, DYS391, DYS393, DYS437, DYS439, DYS441, DYS442, GATA A7.1, GATA H4, G09411; trinucleotides DYS388, DYS392, DYS426; and the dinucleotide repeat YCAII. Primers for the markers DYS385, DYS389, and YCAII target duplicated regions of the Y chromosome and thus can produce two polymorphic peaks with each primer set. The multiplexes we are developing are the first to include all of the European 11-locus "extended haplotype" in a single reaction.

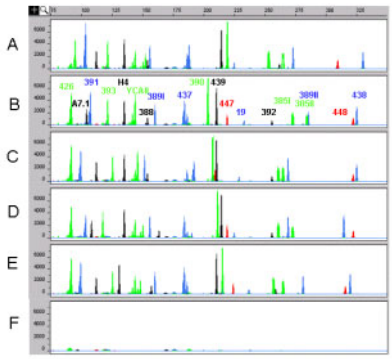
In order to improve the power of discrimination for Y chromosome tests, we have developed strategies for rapidly preparing multiplex PCR assays that utilize both four and five dye chemistries for detection and permit simultaneous amplification of 20 or more Y chromosome STR markers in a single reaction. An important design aspect of our multiplex assays is that PCR product sizes are kept under 350 bp in order to ensure a greater success with testing degraded DNA samples. Primer design issues are considered and efforts to avoid any homology with X chromosome sequences. Primers have been redesigned from previously published work with these Y STR markers in order to make them more compatible in a multiplex amplification. The robustness of the multiplex will be tested in three independent laboratories. In addition, allele ranges for each of the Y STR markers have been well characterized in a diverse set of world population samples.

Purpose: The development of high-level multiplex PCR reactions for typing STRs located on the Y chromosome for human identification purposes.

Major Accomplishments: The Y STR 20 loci multiplex is the largest multiplex PCR system currently available for typing the Y chromosome. Oral presentations discussing megaplex assay development were presented at the 19th Congress of the International Society of Forensic Genetics (Munster, Germany; August 2001) and the 12th International Symposium on Human Identification (Biloxi, MS; October 2001).

Impact: Forensic typing laboratories across the world have expressed an interest in using and evaluating the NIST Y STR multiplex primer set.

Future Plans: Further testing of the 20plex PCR primer set for genetic population studies as well as human identification purposes.


A
B
C
D
E
F
426
391
H4
393
YCAII
389
437
439
A7.1
388
447
19
392
389
389B
389A
438
448

MALDI-TOF Mass Spectrometric Examination and Comparison of the Proteomes from Different Species of Bacteria (whole and disrupted) and Effect of Cell Growth Stage

CSTL Programs: Healthcare Measurements, DNA Technologies
Authors: *K.A. Holland (IPA, Gettysburg College), B.C. Levin; C. Nelson, and B. Nelson (839)*

Abstract: Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is a versatile tool in biochemistry, especially in the area of protein analysis. This approach has been used in the identification of proteins and peptides from protein digests following separation by 2D gel electrophoresis of proteins from whole cells and subcellular organelles. One of the next major endeavors in biology, now that the human genome has been sequenced, is the even more difficult research area of proteomics, namely, deciphering the total protein complement from cells or tissues. In our laboratory, MALDI-TOF MS is being used to directly examine and compare the proteomes of different bacterial species (using both whole and disrupted cells and eliminating the 2D gel electrophoresis step) with the objective of trying to produce cellular spectral 'fingerprints' to identify the various bacteria. We are also examining the effects of different cell growth conditions and cell growth stage on the spectral fingerprints of the whole cell samples. We have demonstrated that freshly grown gram-positive bacteria, *Bacillus subtilis* (strain *B. globigii*), and gram-negative bacteria, *Escherichia coli* (containing the MS2 phage), have unique proteomic MS spectra and that storage of samples at –20^{o}C for two weeks does not affect the spectra. Although the stage of growth (logarithmic, stationary, decay) does significantly change the spectral fingerprint, it appears that there are characteristic identifying signals.

Purpose: There is a need for rapid and efficient methods for 'fingerprinting' various biological samples ranging from bacteria and viruses to proteins and possibly toxic contaminants in air, water and food, to the detection of distinguishing protein biomarkers of disease (e.g., cancerous cells versus non-cancerous cells). MALDI-TOF MS can analyze femtomole (10^{-15}) concentrations of peptides and is the ideal instrument to analyze environmental samples containing very low concentrations of organisms and/or biomolecules. Quick detection and identification of various biological samples without further extensive manipulation, such as PCR, sequencing, protein digestion, and 2D gel electrophoresis, allow for the rapid responses necessary to protect environmental and human health. A database of spectral fingerprints will need to be compiled so that database searching algorithms can provide unambiguous identification of biological samples. In order to establish the usefulness of MALDI-TOF MS for identifying biological samples (whole cell, protein, and viral), we are currently focusing on the proteomic differences between various bacteria and the effects of the growth stage and growth media on the spectral fingerprint of each bacterial species.

Major Accomplishments: As seen in Figure 1, the MS spectra of the gram negative *E. coli* (Fig.1A) and gram positive *B. subtilis* (Fig.1B) are unique and easily distinguishable from each other. Both bacterial species were collected during log phase growth (OD_{600} = 0.972 and 0.816 respectively). The cells were collected as a pellet by centrifugation, washed, resuspended in a matrix solution, dried on a plate and analyzed by MALDI-TOF MS. In addition, we found that *E. coli* and *B. subtilis* samples collected and stored at –20^{o}C for 15 days had spectra nearly identical to the spectra of the samples collected and analyzed immediately (the effect of freezing for longer periods of time is currently being examined). Samples collected at various growth stages (logarithmic, stationary, post-stationary, decay) produced significantly different spectra.

But cultures grown on different days and collected during the log growth stage exhibited very similar spectra.

Impact: Detection of low levels of microorganisms and biomolecules is very important in today's society with our current and mounting concerns regarding biological warfare and contaminated food, soil, air, and water. If we can show that MALDI-TOF MS has the capabilities to lower our level of detection (to the femtomole level) of chemical species, can distinguish between different bacterial species, can detect toxic products in the environment and contaminants in our food, all in a rapid and efficient manner, it will greatly improve our ability to respond to these dangers and hopefully, prevent their damage. The development of a mobile robust instrument would be extremely helpful for immediate responses (detection, identification, clean up, quarantine) in the field to any unknown contaminant.

Future Plans: We are planning to examine multiple different organisms, both prokaryotic and eukaryotic. The lowest level of detection, or the minimum number of cells or protein molecules, will also be determined. In addition, we will be comparing the MS spectra from whole cells and from cell extracts to determine which preparation provides the most distinctive and unique identification biomarkers. We also plan to compare the MS spectra of plasmid-infected bacteria to the same species bacteria without plasmids. If distinct biomarkers can be determined for specific organisms, it may be possible to examine complex environmental samples and identify the individual organisms in these complex mixtures.

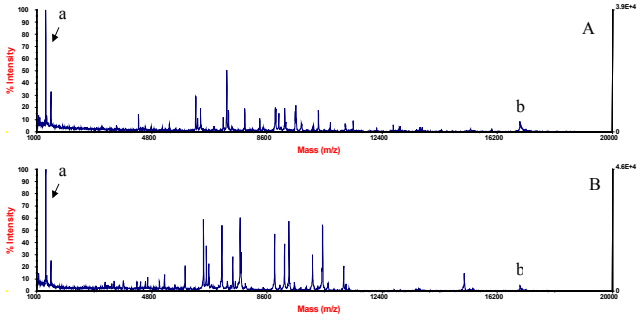


Figure 1: The MALDI-TOF mass spectra for *E. coli* collected at OD_{600} = 0.972 (1A; gram negative) and *B. subtilis* collected at OD_{600} = 0.816 (1B; gram positive). Internal standards are a) angiotensin (MW = 1297.51 g/mol) and b) myoglobin (MW 16952.56 g/mol).

STR Typing with the ABI PRISM™ 3100 16-Capillary Genetic Analyzer

CSTL Program: DNA Technologies
Authors: *J.M. Butler, M.C. Kline, and R. Schoske*

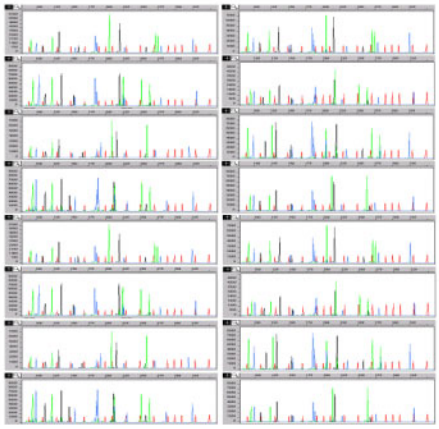
Abstract: We have had a ABI 3100 16-capillary system in our laboratory since April 2001 and have successfully analyzed a variety of STR typing kits including Promega's PowerPlex® 16 and Applied Biosystems' Identifiler™ 16plex kits. These commercial kits and new research multiplex assays we have developed in-house use a variety of fluorescent dye combinations with both 4-dye and 5-dye chemistries. We have generated DNA fragment analysis matrices on the 3100 using various combinations of the following dyes: 5FAM, JOE, NED, ROX, 6FAM, HEX, FL, TMR, CXR, VIC, PET, and LIZ. We have also evaluated performance of samples on both the ABI 310 (single capillary) and ABI 3100 (16-capillary array) instruments.

Purpose: The evaluation of the ABI 3100 16-capillary array instrument for typing STR kits commonly used in the forensic community. Coupling multiplex STR and SNP assays with capillary array systems for high-throughput DNA testing purposes.

Major Accomplishments: Developing successful protocols/validation for typing commercial STR kits on the ABI 3100 and transferring that information to the forensic community. Poster of 3100 data and relevant issues presented at the 12th Annual International Symposium on Human Identification, Biloxi, MS October 2001.

Impact: Assisting the forensic community in validation of the ABI 3100 in order to use instrumentation capable of high-throughput STR typing. These techniques will aid mass disaster investigations and large-scale criminal DNA databasing.

Future Plans: Further use of the ABI 3100 for developing novel Y chromosome STR kits for human identification purposes.



High-throughput Y STR Typing on the ABI 3100 (16-capillary array). The ABI 3100 instrument is capable of producing 256 data points in 45 minutes using new a Y STR 16plex developed at NIST.

Enhancement of the NIST Human Mitochondrial DNA Standard Reference Material 2392 by the Addition of DNA from the HL-60 Cell Line

CSTL Programs: Healthcare Measurements, DNA Technologies
Authors: *K.L. Richie, D.K. Hancock, and B.C. Levin*

Abstract: Mitochondrial DNA (mtDNA) is used by the forensic community for human identification and by the medical community for diagnoses of a number of human mtDNA diseases. A mtDNA Standard Reference Material (SRM 2392) prepared by NIST to provide quality control to the scientific and clinical communities when they amplify and sequence human mtDNA became available in December 1999 (Levin et al., 1999. Genomics 55:135-146). This SRM includes two human DNA templates (CHR and 9947A) and all the information necessary to successfully conduct PCR amplification, cycle sequencing, gel separation, and data analysis to obtain the final DNA sequence. The sequences of 58 unique primer sets, which allow the sequencing of the entire mtDNA (16,569 bp) with no gaps, are also provided. The Federal Bureau of Investigation (FBI) recently requested that NIST enhance SRM 2392 by including DNA from cell line HL-60 which has several evenly spaced polymorphisms in the mtDNA control region and no C-stretch (difficult to sequence) areas. This addition of HL-60 to the NIST human mtDNA sequencing SRM 2392 is currently in progress and should provide additional quality control when amplifying and sequencing human mtDNA. Corroboration of the SRM results will provide assurance that unknown DNA samples are also being amplified and sequenced correctly.

Purpose: The FBI needs DNA SRMs to provide the quality control and assurance that forensic laboratories in the U.S. are sequencing unknown DNA samples correctly. On July 15, 1998, the FBI Director signed Standard 9.5 that stated "The laboratory shall check its DNA procedures annually or whenever substantial changes are made to the protocol(s) against an appropriate and available NIST Standard Reference Material or standard traceable to a NIST standard." The FBI's Combined DNA Index System (CODIS) program now includes mtDNA sequences from unidentified remains, as well as from relatives of missing persons. In order for authorized laboratories to contribute or examine these indices, the FBI has deemed that certain quality standards must be met. In particular, a positive control from the human cell line HL-60 must be run in conjunction with their unknown samples. The primary customers are the forensic laboratories throughout the U.S. and the clinical laboratories that screen the DNA of patients for mitochondrial diseases.

Major Accomplishments: The addition of HL-60 to SRM 2392 is in progress. We will be presenting preliminary results at the 12th International Meeting on Human Identification in October 2001 and at the Annual Meeting of the American Academy of Forensic Sciences in February 2002.

Impact: The inclusion of HL-60 DNA in SRM 2392 will enhance this standard reference material's utility to all the U.S. forensic laboratories that fall under the jurisdiction of the FBI. Since we are providing all the data on the entire mtDNA (all 16,569 bp), it will also provide another positive control for the clinical laboratories that screen patients for mtDNA mutations and diseases and for the toxicologists who screen mtDNA for mutations. It also provides a positive control for any investigator who wants assurance that their laboratory is amplifying and sequencing DNA (nuclear or mitochondrial) correctly.

Future Plans: We plan to amplify and sequence the entire mtDNA of HL-60 at least twice. We will then send the DNA samples and primers to a number of laboratories that agree to participate in an interlaboratory evaluation (ILE). The ILE is conducted to determine and resolve any discrepancies and to assure that all participants obtain the same sequence.

Recovery of DNA from Aged Bloodstains on Untreated Paper

CSTL Program: DNA Technologies
Authors: *M.C. Kline, J.W. Redman, J.M. Butler, and D.L. Duewer (839)*

Abstract: Many reference DNA sample repositories or "DNA banks" are now in existence, primarily for support of epidemiological and genetic research or to enable identification of forensic evidence or human remains. Whole blood, plasma, and buccal epithelium are convenient and minimally intrusive sources of DNA for current DNA analysis technologies. The nature of the sample, how it is collected, and how it is stored are critical issues for the ultimate utility of any DNA banking effort.

Successful DNA typing requires that samples contain an adequate quantity of DNA and that this DNA can be isolated from polymerase chain reaction (PCR) inhibitors (heme, proteins, and many other whole blood components). Current methods of DNA typing use multiplexes of Short Tandem Repeat (STR) loci detected as PCR amplified products ranging in size from 100 through 450 nucleotide basepairs (bp).

We have examined over 300 anonymous bloodstains that have been stored on untreated Schleicher & Schuell 903 paper (S&S 903) from 2 to 15 years at ambient temperatures with no humidity control. As well as samples that were stored at -20°C for 6 years. Our examinations included different methods of extraction (Chelex®, and salting-out) as well as evaluation of the quality of the recovered DNA (yield gel), and typeablity of the DNA obtained.

All samples yielded typeable DNA. A loss of some of the larger STR loci was noted in some of the older and more degraded samples. Images of yield gels indicate that the DNA extracted from the 6 yr old samples stored at -20°C had intact DNA greater than 12 kb in size while the ambient samples appeared as smears of DNA with sizes ranging from 12 kb down to approximately 100 bp.

Purpose: To assess the stability of "field" collected and ambient stored aged bloodstains for STR typeablity. Forensic laboratories have DNA banks of reference samples, mostly stored at -20°C – requiring equipment, energy, and storage space resources. The question "Can I store my samples at ambient temperature?" has been asked on numerous occasions.

Major Accomplishments: We have examined over 300 anonymous bloodstains stored on untreated S&S 903 from 2 to 15 years at ambient temperatures with no humidity control, as well as samples that were stored at -20°C for 6 years. All samples yielded typeable DNA. A loss of some of the larger STR loci was noted in some of the older and more degraded samples. Images of yield gels indicate that the DNA extracted from the 6 yr old samples stored at -20°C had intact DNA greater than 12 kb in size while the ambient samples appeared as smears of DNA with sizes ranging from 12 kb down to approximately 100 bp.

Impact: Our quantitative results enable DNA banks to make a more informed decision on whether or not they should continue to store samples at -20°C.

Future Plans: We intend to continue longitudinal studies on aged samples stored on several media and under a variety of conditions.

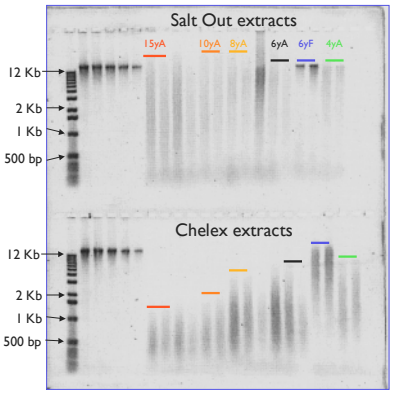
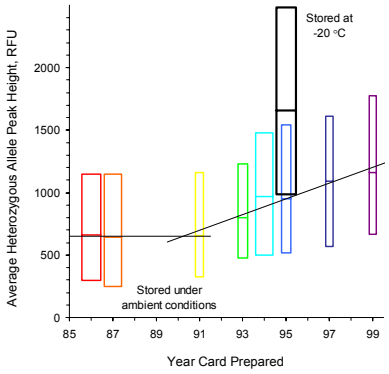


Fig 1. DNA extracted from the 6 year old freezer (6yF)samples have intact DNA greater than 12 kb in size while ambient samples appear as smears of DNA with sizes ranging from 12 kb down to approximately 100 bp.

Fig 2. Median allele height as a function of storage time. Each "box" denotes the central 50% of each height distribution (25%, median, 75%). The box width is proportional to the number of alleles. The lines suggest a possible "plateau" in height loss after about a decade's storage.



High-Throughput Analysis of Telomerase by Capillary Electrophoresis

CSTL Program: DNA Technologies

Authors: *D.H. Atha, and W. E. Highsmith (Department of Pathology, University of Maryland School of Medicine)*

Abstract: The enzyme telomerase has been demonstrated to be expressed in 85-90 % of all human cancers, but not in normal tissues. Clinical assays for telomerase in easily obtained body fluids would have great utility as non-invasive, cost-effective methods for the early detection of cancer. The most commonly used method for the detection and quantification of telomerase enzyme activity is the PCR based assay known as the telomerase repeat amplification protocol or TRAP assay. Most of the TRAP assay systems use a slab-gel based electrophoresis system to size and quantify the PCR-amplified extension products. We are developing high-throughput capillary electrophoresis (CE) methods for the analysis of TRAP/PCR products. TRAP/PCR products were generated from cancer cell lysates using fluorescent - labeled primers and analyzed on the Applied Biosystems Model 310 capillary electrophoresis system. Results were compared with unlabeled TRAP/PCR products analyzed on the Bio-Rad BioFocus 3000 capillary electrophoresis system using SYBR™ Green I dye. Both CE systems were able to resolve the TRAP ladder products with high reproducibility and sensitivity.

Purpose: Telomerase activity measurements are of great interest to the medical community for the diagnosis of cancer. These tests would have particular utility as non-invasive, cost-effective methods for the early detection of cancer. To be effective, the methods must be optimized for sensitivity, reliability and throughput. Currently, telomerase assays measure the small amount of enzyme activity in tissue lysates or cells collected from body fluids. The most commonly used system for the detection and quantification of telomerase activity is the PCR based assay known as the telomerase repeat amplification protocol or TRAP assay. Most of these assay systems use a slab-gel based electrophoresis system to size and quantify the PCR products. We are developing more efficient methods for analyzing these PCR products using capillary electrophoresis (CE). The method will streamline validation of the telomerase TRAP assay system with increased sensitivity, reproducibility and automation over slab-gel methods.

Major Accomplishments: We have developed two capillary electrophoresis methods for the analysis of telomerase. The TRAP assay was conducted on lysates of the human lung cancer cell line A-549 in reactions containing 2 to 500 cells. For the first method, TRAP/PCR products were generated using fluorescent (5-carboxyfluorescein) - labeled CX and TS primers and analyzed on the Applied Biosystems Model 310 capillary electrophoresis system using POP4™ polymer. After analysis with GeneScan™ and Genotyper™ software, the total peak area of the TRAP ladder extension products were computed using Microsoft Excel™ . For the second method, unlabeled TRAP/PCR products were analyzed on the Bio-Rad BioFocus 3000 capillary electrophoresis system using 6% HMW PVP (high molecular weight polyvinylpyrrolidone) polymer and SYBR™ Green I dye. Both CE systems were able to resolve the TRAP ladder products with high reproducibility and sensitivity (5 cells). Results of this work have been presented to the Early Detection Research Network of the National Cancer Institute and at the 12th Annual Frederick Conference on Capillary Electrophoresis at NCI Frederick.

Impact: An accurate, reproducible and efficient system to measure telomerase in clinical samples will be required for its validation as a prognostic indicator and monitor of cancer. High-throughput methods will be necessary for the analysis of the large numbers of samples required

for validation of the telomerase assay. With the appropriate robotic sample handling system, this CE method, as described above, would streamline validation of the telomerase TRAP assay with increased sensitivity, reproducibility and automation over slab-gel methods. This method would also enable accurate analysis of the large number of clinical samples required for validation of telomerase assays for the early detection of cancer.

Future Plans: The next phase of the project will involve the application of the CE methods to high-throughput analysis. This will involve the adaptation of the current TRAP procedure to commercially available robotic systems, multi-capillary analysis on the ABI 3100 CE and high-throughput data analysis using GeneScan, Genotyper and Excel.

NIST-EDRN Biomarker Validation Laboratory, Workshop on Nanotechnology and Early Cancer Detection

CSTL Program: DNA Technologies

Authors: *P.E. Barker, and P. Srinivas (NCI)*

Abstract: As part of the ongoing ***NIST-EDRN Biomarker Validation Project***, we organized a NIST-NCI Workshop on August 30 and 31, 2001, focused on *early cancer detection and nanotechnology*. To promote discovery, validation and clinical implementation of novel cancer biomarkers, new technology and interdisciplinary research will play an important role. To decrease costs, high throughput, miniaturized assays would be ideal. Therefore, a workshop featuring 20 leaders in biologically applicable nanotechnology was held at NIST. One hundred fifty registrants attended from academia (25%), government (60%) and the commercial sector (15%). Co-Chairs were Dr. George Whitesides and Dr. Lee Hood. The meeting was financially supported by the NCI, by the NIST ATP Program and by the DNA Technologies Group.

Purpose: Nanotechnology and cancer researchers rarely interact. These groups have fundamentally different backgrounds and bring to problems a very different approach. Because they attend different research meetings, we considered it worthwhile to bring together these two scientific and technology communities to summarize opportunities and technology from emerging areas of nanotechnology that may impact on early detection of human cancers. Further, the meeting elaborated a role for NIST in this endeavor.

Major Accomplishments: To date, the overall accomplishments of the NIST Biomarkers Validation Laboratory follow. 1) Initiation of validation analysis on the technology of fluorescence in situ hybridization for cancer risk analysis and completion of the first phase of the study. This includes analysis of breakage rates in 2400 metaphase cells with dual color FISH probes for breakage on human chromosome 5, thought to be an index of cancer susceptibility. Initial data is currently undergoing statistical analysis to determine inter-laboratory and inter-observer variability. 2) Initiation of technical improvements ("pre-validation") on PCR/DNA sequencing technology for analysis of mitochondrial DNA base mutations as a tool for early detection of lung cancer. Progress here has been re-evaluation of previously published primers, simplifying and streamlining PCR products synthesis, and improvements in the ability to PCR amplify mtDNA from diverse racial classes of samples reflecting diversity in U. S. cancer patient populations. 3) Initiation of application of capillary electrophoresis for analysis and quantification of the enzyme telomerase for early cancer detection ("pre-validation"). Here the CE methods improve reproducibility and sensitivity compared with slab gel analysis of telomerase-generated "TRAP-assay" PCR products. (See Atha and O'Connell highlights for details).

In addition to lab work at NIST, ***the Nanotechnology and Early Cancer Detection Workshop***, the first of several planned NIST-NCI meeting collaborations, brought many scientists and businessmen to NIST, introducing them to colleagues in other fields and to the institution. In a meeting report to be published in *Nature Biotechnology* (writing in progress), a series of areas was identified where nanotechnology may have the greatest impact on cancer detection. In addition, the meeting concluded with a list of recommendations to NIST indicating where the Institute can have the greatest impact in fostering nanotechnology applications and standards as applied to health care and cancer diagnostics.

Impact: The Workshop introduced both cancer biologists and nanotechnologists to NIST facilities and programs. At least one workshop speaker involved in a biotech company who was unaware of NIST programs, has applied for support as the immediate, quantifiable result of this workshop. The Workshop thus provided a direct interface between nanotechnology developers and the NIST ATP Program. There were representatives from as far away as Cambridge (UK), Tokyo and Paris and from a range of U. S. companies such as Intel and Hoffmann-LaRoche. Thus, in addition to success at encouraging scientific interactions, the meeting was an opportunity to showcase NIST activities in nanotechnology and healthcare.

Future Plans: The Nanotechnology Workshop at NIST was taped by a cable TV company for a science news story as well audio taped for direct Internet broadcast. The NIST Public Relations department was asked whether NIST could help with a future cable TV program devoted nanotechnology. This is under consideration at present. A second NIST-NCI Workshop is in the planning stages focused on development of a SNP Database for Early Cancer Detection Genes. We will continue our laboratory validation work (chromosomes, mtDNA and telomerase). As these are completed, we anticipate new validation projects as needs arise.

Proteomics, Structure Genomics - Elucidation of function of hypothetical proteins revealed from genome sequencing projects

CSTL Program: DNA Technologies
Authors: *P. Reddy, J. Moult, J. Orban, O. Herzberg, and G.L. Gilliland*

Abstract: We have identified a number of genes for hypothetical proteins from *Haemophilus influenzae*. We cloned about 50 hypothetical genes into the protein expression vectors, overproduced the proteins. About 50% of the overproduced proteins were soluble which were purified to homogeneity. The other 50% formed inclusion bodies which are essentially protein aggregates and could not be used further. The crystal structure of 18 proteins was solved.

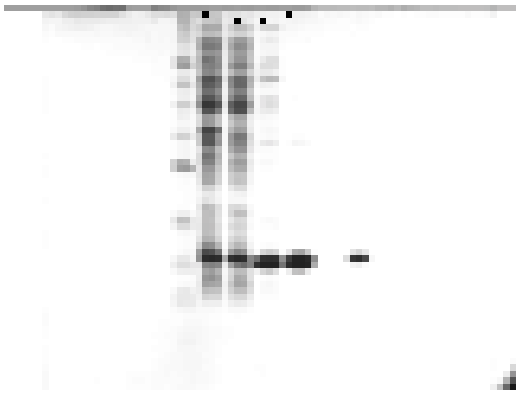
Purpose: To clone hypothetical genes from *Haemophilus influenzae* genome, express and purify the corresponding protein, and determine the three dimensional structure of the proteins by X-Ray crystallography to assign a possible function for the protein. As the genomes of the microbial and eukaryotic origin are being decoded, we are learning how little we know about the function of numerous proteins in the cell. Hence, this project is aimed at unraveling the function of the proteins with no known function, the so-called hypothetical proteins.

Major Accomplishments: Proteins expressed in the pRE system were purified usually in two steps. The first step was either anionic or cationic exchange based on the pI of the protein. Proteins were further purified on gel permeation chromatography. These two protocols yielded nearly homogeneous proteins for well expressed proteins (>20% of the total cellular protein). For proteins expressed in the 5-15% range, the purity of the protein was about 90-95%. Proteins expressed with histidine tag in the pET system were purified on Nickel affinity resin in a single step and cleaved with thrombin to generate the native protein. The histidine tag was separated from the native protein by passage through the Nickel affinity resin a second time where the histidine tag is bound and the native protein simply passes through the column.Pure proteins were crystallized and the X-Ray diffraction pattern of the crystals was collected. Structure of 18 proteins was solved.

Purification of a hypothetical protein:

From left to right

Lane 1: Molecular weight marker
Lane 2: Crude cell extract
Lane 3: Soluble extract
Lane 4: Ion exchange chromatography
Lane 5: Gel permeation chromatography



Impact: Innumerable genome sequences of microbial, archea, and eukaryotic origin are being decoded. Consequently, we are uncovering many proteins of unknown function termed as hypothetical proteins. Development of protein pharmaceuticals and vaccines for infectious diseases by the industry is being hampered by the lack of knowledge about the hypothetical proteins. The structure genomics endeavor will help the pharmaceutical industry to advance the

development of drugs for infectious diseases, inherited diseases, and other wide spectrum of diseases.

Future Plans: We will continue to work on the structure / function elucidation of all the hypothetical proteins (about 500) from a microbial origin, *Haemophilus influenzae*.

Standardization of PCR Amplification for Fragile X Trinucleotide Repeat Measurements

CSTL Program: DNA Technologies
Authors: *K.L. Richie, D.H. Atha, J.P. Jakupciak, and C.D. O'Connell*

Abstract: To provide the clinical diagnostics community with accurate protocols and measurements for the detection of genetic disorders, we have established a quantitative measurement program for trinucleotide repeats associated with human disease. In this study, we have focused on the triplet repeat associated with fragile X syndrome. Five cell lines obtained from the Coriell Cell Repository were analyzed after PCR amplification and size separation. These cell lines were reported to contain CGG repeat elements (ranging from 29 to 110 repeats). Our initial study focused on measurement variability; a) between slab gel (SG) and capillary separation systems, b) variability associated with amplification, c) inter-lane variability (SG), and d) inter-gel variability (SG). Samples were run in triplicate for all measurements, and the analysis was performed using GeneScan ™ software. The repeat sizes were verified by DNA sequence analyses. The standard deviations for inter-lane measurements in slab-gels ranged from 0.05 to 0.35. There was also little variation in size measurements performed on different gels and among PCR amplifications. The CGG repeats measured by capillary electrophoresis (CE) were even more precise, with standard deviations ranging from 0.02 to 0.29. The slab gel and CE size measurements were in agreement for normal alleles, however the premutation alleles, yielded significantly smaller sizes by CE.

Purpose: CSTL began a DNA measurement program in 1989 to address standardization needs for human identification and began a DNA diagnostics program in 1993. This project is the first in the field of human genetics. As genetic information becomes more widely available largely due to the Human Genome Project, the causes of an estimated 6,000 genetic diseases will become linked to specific genetic variants. Fragile X syndrome is one of these diseases, which belongs to a broader group of 14 other human diseases that are caused by the expansion of a triplet repeat sequence. A critical problem for clinical diagnostics of fragile X syndrome is accurate CGG repeat size measurements in the FMR1 gene. In general, the larger the number of CGG repeats, the greater the disease severity. Unreliable size determinations create poor information for genetic counseling and pre-natal screening purposes. This research is expected to result in optimized laboratory protocols to provide accurate measurement capabilities for triplet repeats. The primary customers are the clinical diagnostics community and those genetic testing laboratories that screen the DNA of patients for triplet repeat diseases; specifically, fragile X syndrome.

Major Accomplishments: The results of this work have been presented at the 2001 Clinical Genetics Meeting (Miami, FL), the International Conference of Human Genetics (Vienna, Austria), and at the American Society of Human Genetics (San Diego, CA) in October, 2001. Research results have been submitted to *Clinical Genetics* and accepted for publication.

Impact: Fragile X is the leading heritable cause of mental retardation; thus, accurate quantitation of the number of repeat sequences is necessary for carrier screening in family planning as well as for accurate diagnosis of disease and pre-natal screening. Therefore, we feel that our research will have the largest impact on diagnostic testing by focusing on the accurate quantitation and standardization of Fragile X triplet repeat measurements.

Future Plans: Our future plans include the development of a Fragile X syndrome Standard Reference Material to provide accurate size determinations of CGG repeat length mutations. We will perform a limited interlaboratory evaluation with selected laboratories to assess state-of-the-art CGG repeat size measurement methodology. A NIST SRM composed of FMR1 size standards issued with a certificate of analysis that reports the results of their characterization will ensure the accuracy and comparability of measurement results among laboratories. This SRM would be primarily used for clinical diagnostics but could also be useful for population genetics.

Participation in a Consortium to Evaluate an Automated Microsphere Suspension System to Detect Multiplexed Single Nucleotide Polymorphisms (SNPs) in Human Mitochondrial DNA

CSTL Programs: Healthcare Measurements, DNA Technologies
Authors: *D.K. Hancock, and B.C. Levin*

Abstract: A consortium consisting of forensic, industrial, academic, and government laboratories is evaluating the capability of an automated microsphere suspension system to detect multiplexed SNPs in both human mitochondrial DNA (mtDNA) and the Y-chromosome. We were invited to participate in this consortium because of our mtDNA expertise. The Luminex 100 system can conduct up to 100 concurrent bioassays on a single sample using arrays of up to 100 different reactive microspheres identified by precise ratios of spectrally distinct internal fluorochromes. A third reporter fluorochrome coupled to the sample molecule quantifies the extent of biomolecular interaction between the sample and the microsphere surface. Microspheres are individually examined in a rapidly flowing fluid stream as they pass between two laser beams. The autosampler provides versatile, high-throughput and multiplexing capabilities. Using a 96-well microtiter plate with a 100 different microsphere-bound analytes per well allows 9600 assays per plate. Thus, DNA and RNA SNPs as well as proteins can be analyzed depending on the analytes (e.g., oligonucleotides, antibodies, antigens, or enzyme substrates) bound to the microspheres. All participants in the consortium received 2-3 days of training and examined mtDNA fluorescently labeled PCR amplicons from two different DNA templates in triplicate. These amplicons were hybridized to allele-specific oligonucleotides bound to a microsphere suspension array.

Purpose: Now that the human genome has been sequenced, investigators are examining the multiple SNPs, mutations, insertions and deletions found in populations to determine those responsible for the diversity of the population and those associated with diseases. Other scientists are investigating the proteins that are formed from the 30,000 genes now believed to compose the human genome and are trying to determine the function, the interactions and the diversity of the human proteome. This automated liquid bead system has the potential to examine multiplexed DNA, RNA and protein samples. With its 96 well plate, it has the capability of examining up to 96 samples, each of which can contain multiplexed areas for analysis. The customers are the scientists trying to determine the SNPs responsible for population diversity and the health care industry trying to correlate mutations with diseases as well as the forensic community who will use the mtDNA and Y-chromosome kits for human identification.

Major Accomplishments: We have completed the training and examination of the mtDNA samples from two different cell lines and two different mtDNA areas. Analysis of the data indicated that the instrument worked well in detecting the presence of specific mtDNA SNPs. The results of the consortium will be presented at the 12th International Meeting on Human Identification in October 2001 and at the annual meeting of the American Academy of Forensic Sciences in February 2002.

Impact: At the present time, the task of examining individually all the SNPs and disease-correlated mutations in human populations is daunting. This methodology provides the ability to detect specific SNPs and mutations in a multiplex format and in multiple samples and will greatly enhance our ability to accomplish these goals in a high-throughput manner. The work on

the mtDNA and Y-chromosome will also be useful for human identification by the forensic community. The ability to find specific proteins in the complex mixtures that comprise the proteome or environmental samples will also enhance our ability to detect specific proteins of concern and to determine their function.

Future Plans: We plan to investigate the sensitivity of this methodology by examining our heteroplasmic mtDNA Standard Reference Material 2394 that consists of an amplified region containing a single base substitution at concentrations of 1, 2.5, 5, 10, 20, 30, 40, and 50%. We also plan to investigate the capability of the instrument to detect specific proteins in complex protein or environmental mixtures.

Prototype Y Chromosome Standard (SRM 2395)

CSTL Program: DNA Technologies
Authors: *J.M. Butler, M.C. Kline, and R. Schoske*

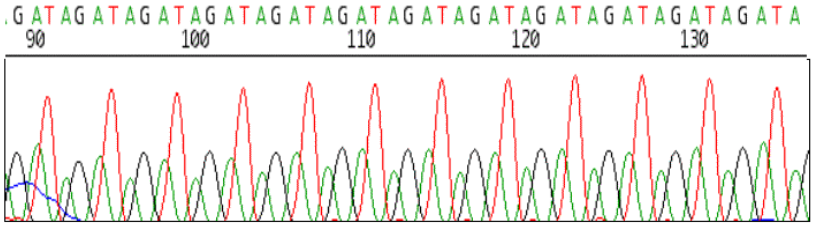
Abstract: In an effort to support the growing demand for Y chromosome testing, the National Institute of Standards and Technology (NIST) is developing a standard reference material (SRM) that can be used to calibrate instrumentation and verify assay performance with Y STR and Y SNP markers. Over 140 candidate DNA materials for SRM 2395 have been screened. From these samples, five male and one female DNA samples have been selected and extensively characterized. These candidate materials reflect multiple alleles at each Y STR and SNP locus. We are testing these six candidate (A thru F) SRM samples with commercially available and research multiplex assays. In addition, all of the DNA samples are being sequenced to confirm exact repeat compositions of the STR markers. SNP markers are being analyzed with multiple technologies to confirm the polymorphic nucleotide present in each sample.

Purpose: Providing a Y chromosome standard reference material for calibration of DNA typing instrumentation.

Major Accomplishments: The SRM is currently in prototype phase. Candidate SRM samples have been selected and characterization is currently underway. Poster of SRM prototype data presented at the 12th International Symposium on Human Identification Biloxi, MS October 2001.

Impact: Will be used by the forensic and paternity testing communities for the standardization of Y chromosome assays and well as instrumentation performance.

Future Plans: Completion of SRM 2395 and assessing its utility to the forensic community.



Sequencing in POP-4 on ABI 310 of Candidate D

Electropherogram of sequencing results from SRM candidate D. The sequencing data confirms the presence of 12 short tandem repeat units (GATA) in the DNA sample.

Design and Use of a Peptide Nucleic Acid for Detecting the Heteroplasmic Low-Frequency Mitochondrial DNA Mutation Associated with the Disease MELAS

CSTL Programs: Healthcare Measurements, DNA Technologies
Authors: *D.K. Hancock, F.P. Schwarz, B.C. Levin, F. Song (CARB/UMBI), and L.-J. Wong (Georgetown University Medical Center)*

Abstract: A multitude of mitochondrial DNA (mtDNA) diseases have been correlated with single nucleotide polymorphisms (SNPs), mutations, insertions, and deletions. Most of these diseases are neuromuscular, but deafness, diabetes, epilepsy, progressive dementia, hypoventilation, cardiac insufficiency, renal dysfunction, and sudden onset blindness are some of the other symptoms of mtDNA mutations. Most pathogenic human mtDNA mutations are heteroplasmic (i.e., the mutant mtDNA coexists with the normal mtDNA in the tissue or cell). The mutation level varies with the tissue and is often difficult to detect (especially in blood samples) when very low levels of the mutant exist in a population of normal mtDNA molecules. We are using a simple methodology to detect low levels of the single base pair heteroplasmic MELAS (Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-like Episodes) (A3243G) mutation. A series of peptide nucleic acids (PNAs) was designed to bind to the wild type mtDNA, thus reducing the polymerase chain reaction (PCR) amplification of the wild type mtDNA to background levels while permitting the mutant DNA to become the dominant product and readily discernable. This methodology permits detection of the MELAS A3243G mutation in asymptomatic or symptomatic carriers with low to undetectable blood levels of this mutation.

Purpose: This work addresses the problem of detecting low-frequency single mtDNA mutations in order to help in the diagnosis of mitochondrial diseases and provide more predictive genetic counseling and assistance in screening of *in vitro* fertilized embryos to prevent the inheritance of these devastating mitochondrial diseases. It could also be useful in the screening of mixed population samples for low-frequency SNPs. Now that the human genome has been sequenced, the next big effort is to detect the SNPs and mutations, discover their effects and determine their population frequency.

Major Accomplishments: In the past year, we found a source (Georgetown University Medical Center) of DNA from MELAS patients. We designed, synthesized, characterized and purified a series of PNAs and tested them to determine the best one and the optimal conditions for blocking the amplification of the wild-type mtDNA while allowing the mutant DNA to amplify. We also examined the effect of the PNA and DNA concentrations. With our optimized PCR conditions, we were able to show that the mutation that was not visible or barely visible in the sequence electropherogram in the absence of PNA became the dominant peak and readily detectable in the presence of our designer PNA. These results were reproducible with DNA from eight different MELAS patients. We presented preliminary results at the Biotechnology Offsite in April 2001 and will present additional results at the 12th International Meeting on Human Identification in October 2001. We are currently preparing a manuscript for publication in a peer-reviewed journal.

Impact: The A3243G mtDNA point mutation is the second most common mtDNA defect. In addition to its association with MELAS, it is the most common mtDNA mutation associated with diabetes mellitus. When these results are published and when a SRM is available, there will be a considerable impact on the clinical community who screen patients for mitochondrial diseases. It

will greatly facilitate the detection of a mutation that was previously extremely difficult to detect and will help to prevent false negative results. It will provide help to those doing genetic counseling and those screening human oocytes for *in vitro* fertilization processes. It will also show the proof of principle for scientists screening populations for specific SNPs.

Future Plans: We plan to develop a NIST SRM for the detection of the MELAS mutation in human mtDNA. This SRM will consist of the specially designed PNA, the primers needed to amplify the region where the mutation occurs, and the optimized protocol to allow any laboratory with PCR and sequencing capabilities to use this SRM and detect the MELAS mutation in blood even if it is not detectable in the absence of the PNA. Since many of the human mtDNA diseases have been correlated with single base substitutions, we hope to add more specific PNAs and primers to this SRM in the future to detect additional human mtDNA mutations. We also plan to extend this work to study the prevalence of the A3243G mtDNA mutation in diabetes mellitus patients.

MitoAnalyzer, a Computer Program and Interactive Web Site to Determine the Effects of Single Nucleotide Polymorphisms (SNPs) and Mutations in Human Mitochondrial DNA

CSTL Programs: Healthcare Measurements, DNA Technologies
Authors: *B.C. Levin, and M.S. Lee*

Abstract: MitoAnalyzer, an interactive human mitochondrial DNA (mtDNA) web site to enable investigators to determine the effects of any single polymorphism, mutation, insertion or deletion in human mtDNA has been developed and went public March 16, 2001. Using the same numbering system as that used by the Cambridge Reference Sequence (1), the investigator enters the number and the identity (A, C, G, T) of the affected base. The program compares that to the Cambridge Reference Sequence and will provide information on where the change occurs (i.e., whether it is in the coding or non-coding region, the HV1 or HV2 region, in the first, second or third bp of the codon, whether it affects a ribosomal RNA, a transfer RNA, or a messenger RNA coding for a protein), which protein is affected, and whether it causes a change in an amino acid in that protein. If an amino acid has changed, it will specify the position of that amino acid change in the protein (e.g., amino acid # 25 in a protein containing 200 amino acids). This program will also provide the entire amino acid sequence of the protein as described in the Cambridge Reference Sequence and the new changed version. Mutations associated with published mitochondrial diseases are noted. This program, thus, facilitates rapid analysis and evaluation of SNPs and mutations in human mtDNA. This web page can be accessed at: http://www.cstl.nist.gov/biotech/strbase/mitoanalyzer.html

Purpose: To develop a computer program and interactive web site to provide information on the location and effects of single nucleotide polymorphisms (SNPs), base substitutions, mutations, insertions or deletions anywhere in the 16,569 base pairs that comprise human mitochondrial DNA (mtDNA) and how the change affects the resulting proteins and whether the change has been correlated with any mtDNA disease. The need for this type of program was recognized after the development of the human mtDNA sequencing SRM 2392 (2) which indicated that the mtDNA had many polymorphisms and/or mutations, the meaning of which were deciphered with great difficulty. At the present time, the human genome is being sequenced to determine the location of SNPs and mutations and the determination of the effects of these SNPs and mutations is difficult and time consuming. We hope that our interactive web site will serve as a model for future web sites for the genes found in the human nucleus. At the present time, the primary customers are the investigators who are sequencing mtDNA for medical diagnoses, for evolutionary purposes, and for mutation detection.

Major Accomplishments: The web site went public on March 16, 2001 and has had 1858 hits since that time. We have presented the work at the 40th Annual Meeting of the Society of Toxicology, San Francisco, CA, March 25-29, 2001, the FDA Genomics/Proteomics Working Group, FDA Parklawn Building, Rockville, MD, May 1, 2001, and the 2nd Annual Biotechnology Division Retreat, Hilltop Hotel, Harpers Ferry, West Virginia, April 19, 2001 and we will be presenting it at 12th International Symposium on Human Identification, Biloxi, Mississippi, October 9 - 12, 2001. The paper has been accepted for publication in the peer-reviewed journal Mitochondrion and is currently in press (3).
Impact: The web site has had 1858 hits since going public March 16, 2001. We have had many favorable comments and a number of other web sites have provided a link to it. We are hoping that this approach will become a model for nuclear genes.

Future Plans: We will be continuing to update the web page as necessary. We hope to copyright the program used to generate the data.

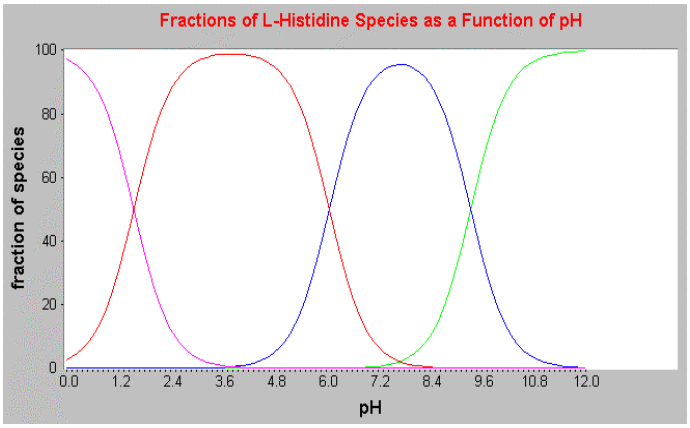
Biothermodynamics

CSTL Program: Chemical and Biochemical Data
Authors: *R.N. Goldberg, and Y. Tewari*

Abstract: A review that contains selected values of thermodynamic quantities for the ionization reactions of 64 buffers, which are used in biological research, has been completed. Since the aim is to be able to predict values of the ionization constant at temperatures not too far from ambient, the thermodynamic quantities which are tabulated are the pK, standard molar Gibbs energy $\Delta_r G°$, standard molar enthalpy $\Delta_r H°$, and standard molar heat-capacity change $\Delta_r C_p°$ for each of the ionization reactions at 298.15 K. The selection of thermodynamic data for each buffer is discussed. A summary data table is available for calculating fractions of the various buffer species present in solution as a function of pH. This has been done for L-histidine in the attached figure.

Purpose: To create a comprehensive tabulation of thermodynamic quantities for the ionization reactions of biological buffers.

Major Accomplishments: A review that contains selected values of thermodynamic quantities for the ionization reactions of 64 buffers, which are used in biological research, has been completed. A summary data table is available for calculating fractions of the various buffer species present in solution as a function of pH. This has been done for L-histidine in the attached figure.



The amino acid L-histidine can form 4 distinct ionic species as a result of changing pH. Using values of thermodynamic properties assembled in the new Biological Buffer Database the figure shows the relative fraction of each species as a function of pH:
purple – L-histidine^{2+},
brown – L-histidine^{+},
blue – L-histidine^{0} ,
green – L-histidine^{-}

Impact: There does not exist a comprehensive tabulation of thermodynamic quantities for the ionization reactions of biological buffers. These buffers are essential components of every biological preparation ranging from nucleic acids through to stem cells and whole tissues. The International Union of Pure and Applied Chemistry (IUPAC) Commissions on Biophysical Chemistry and Thermodynamics identified the need for a comprehensive, critically evaluated tabulation of this data as a critical need in biochemistry.

Future Plans: Preparation of major publications for the archival literature is in progress. A summary table of this database has been invited for inclusion in the next addition of the CRC

Handbook of Chemistry and Physics. In addition, specific measurement needs, including key biochemical substances such as essential amino acids like alanine, have been identified, and NIST staff should be able to do these if sufficient resources can be found.

BioSpectroscopy

CSTL Program: Bio-Molecules and Materials
Authors: *A.K. Gaigalas, L.L. Wang (William and Mary College), and A. Schwartz (Center for Quantitative Cytometry)*

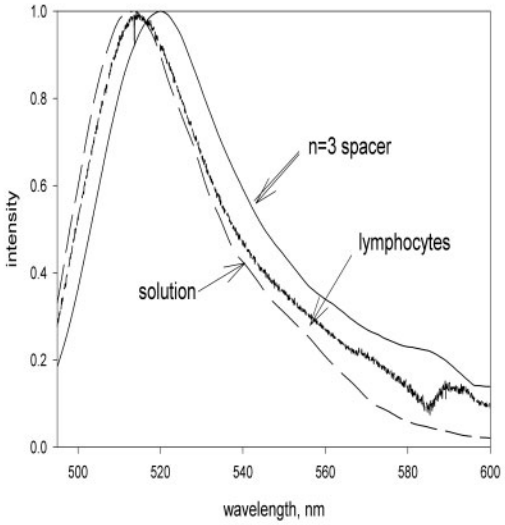
Abstract: In the last year, the first reference material (SRM 1932) for use as a standard in measuring fluorescence intensity has been developed. However, the quantitation of the fluorescence intensity signal is complicated by the sensitivity of fluorescence properties of the fluorophore to the microenvironment. This sensitivity has resulted in difficulty in comparing measurements performed in different laboratories and on different instruments. For example, in flow cytometers the objective is to quantify the number of fluorophore labeled antibodies attached to a given cell population. The cytometer response is calibrated using a set of microbeads each with a different amount of immobilized fluorophore. A calibration curve is generated by plotting the cytometer response as a function of the known amount of fluorophore on the microbead. Current efforts are focused on developing a particle fluorescence intensity standard, SRM 1933. These reference particles will be useful for constructing calibration curves that take into account the difference in the microenvironment of the fluorophore on the microbead and on the cell since the spectral properties of the fluorophore will be different on the two surfaces. For best analytical results the spectral properties of the calibrator and the analyte should be as similar as possible. We have examined the fluorescence emission spectra of fluorescein isothiocyanate (FITC) immobilized on microbeads using spacers of different length.

Purpose: The objective of the work is to identify general principles that determine the fluorescence properties of fluorophores at microbead interfaces, and to use these principles to guide the choice of the immobilization procedures that would yield the best spectral match to the analyte.

Major Accomplishment: The figure shows the comparison of the fluorescence emission spectra of a fluorescein solution (dashed line), a preparation of FITC labeled whole blood (curly line), and the FITC labeled microbead with a 3-carbon spacer (solid line). The whole blood preparation and the microbead suspension are the same as those used in a flow cytometer assay. The microbead suspension is used to calibrate the flow cytometer. The discrepancy between the spectra of microbead and whole blood suggest that the resulting calibration will have systematic errors. The systematic errors would be evident in mismatch of measurements from different instruments that sample different portions of the spectrum. The microbeads with n=12 carbon spacer have the closes match of the excitation and emission spectra and thus are the beast calibrators for this assay. Attention to the cause of spectral shifts may lead to the design of fluorescence intensity standards that yield superior quantitation.

Impact: Fluorescence measurements of various kinds are leading the revolution to quantify biological information. Fluorescence-labeled flow cytometry is used for cellular phenotyping for diagnosing genetic diseases. Two-color fluorescence DNA microarrays are increasingly used to detect diseases arising from multiple gene disorders, and also in the discovery of novel genes, proteins, and metabolic pathways from all life forms.

Future Plans: A long-term goal is to apply SRM 1933 to standardization of fluorescence intensity measurements from flat plate DNA microarray format.



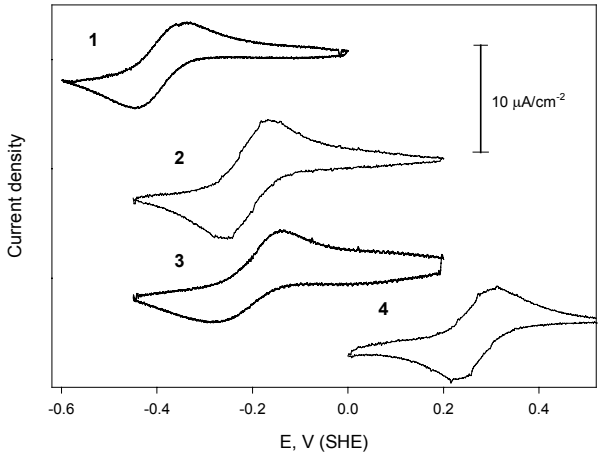
BioElectrochemistry

CSTL Program: Bio-Molecules and Materials
Authors: *V. Reipa, M.P. Mayhew, V.L.Vilker, and G. Valincius (Vilnius State University, Lithuania)*

Abstract: In this work, we have developed fast and sustainable electron transfer to redox proteins on semiconducting metal oxide electrodes by controlling the electronic properties, namely flat band potential and charge carrier concentration. We were able to achieve high electron transfer rates toward two ferredoxins and Cyt c on degenerate metal oxide electrodes without resorting to surface modification. Doping levels and the position of the flat band potential were controlled through the annealing procedure. These CTO electrodes exhibited the highest reported ET rates for the ferredoxins Sp fd and Pdx. The charge transfer properties of these electrode materials can be varied while maintaining their inherent hydrophilicity, a necessary property to prevent electrode fouling and protein denaturation. In addition, these cadmium tin oxide electrodes are transparent in a wide optical range and exhibit electrochemical and chemical stability in the physiological pH range.

Purpose: The direct, facile and sustainable electron exchange between redox proteins and electrodes is of utmost practical significance for the development of biosensors and electroenzymatic synthesis processes. Based primarily on studies with cytochrome c (cyt c), the highest initial electron transfer rates have generally been observed on clean metal electrodes. However, due to rapid protein denaturation on contact with metals, and the propensity of metal surfaces to adsorb organic contaminants, the electron exchange rates decay rapidly unless special electrode surface modification procedures are undertaken to increase process sustainability and/or rate. Our purpose was to adapt hydrophilic metal oxide semiconductor surfaces, which are inherently less susceptible to irreversible adsorption by organics, for electroactivity toward soluble redox proteins.

Major Accomplishment: We were able to achieve high electron transfer rates toward two ferredoxins and Cyt c on degenerate metal oxide electrodes without resorting to surface modification. Doping levels and the position of the flat band potential were controlled through the annealing procedure. These CTO electrodes exhibited the highest reported ET rates for the ferredoxins Sp fd and Pdx.



Cyclic voltammetry (CV) dependences of several redox proteins measured on sol-gel cadmium tin oxide electrodes. 1- spinach ferredoxin (Sp fd) , 2 -putidaredoxin (Pdx) on electrode 1, 3 – Pdx on electrode 2, and 4 – cyt c. CVs measured in Tris buffer with purified proteins (200 μM) using an EG&G potentiostat (Model 173, EG&G Instruments, NJ) at 10 mV/s. The curves are shifted vertically for clarity.

Impact: Rapid, sustainable electron transfer to several kinds of metalloproteins with CTO electrodes will significantly enhance the possibilities for using spectroscopy to study the properties of these proteins, and for using them as transducing agents in biosensors.

Future Plans: Future improvements ot the CTO electrode materials are not envisioned. The CTO electrodes will find numerous uses in our measurements of the properties (e.g. CD and Raman spectra) of several biomacromolecules, and the formal potential of nucleotides and DNA.

Preparative DNA Separations

CSTL Program: Bio-Molecules and Materials
Authors: *K. Cole, M. Markström, and B. Åkerman (Chalmers University of Technology, Sweden)*

Abstract: We have studied the electrophoretic migration of double-stranded T4 DNA (164 kilobasepairs) in gellan gels by velocity measurements and linear dichroism spectroscopy studies of the DNA coil conformation during the migration. The gels either contained 0.3% of high molecular weight ($5 \cdot 10^6$) poly(ethylene oxide) (PEO), in order to suppress the electroosmosis induced by the negative charges of the gellan polymer, or were free of such added polymer and therefore exhibited an electroosmotic flow which is opposite to the DNA migration. In both cases does the viral DNA migrate in an oscillatory manner between stretched and coiled states, because it becomes entangled with the gellan gel fibers. In the stretched state of the cycle the molecules are substantially aligned with the field. As the field is turned off the alignment relaxes first by a rapid (seconds) destretching along the path in the gel, followed by a slower (minutes) end-on type of motion to the equilibrium isotropic coil state. The well-understood migration and relaxation behavior was exploited to investigate the effect of the electroosmotic flow on the DNA migration, and the effect of the PEO added to quench this retarding flow. It was found that the electroosmotic flow strongly reduces the electrophoretic stretching of the DNA, but has small effects on the path chosen by the DNA through the gel. The added PEO has two effects on the DNA migration. An indirect effect is that the quenched electroosmosis leads to a stronger stretching of the DNA, and to shorter cycle period times because the ends of the molecules move faster in the absence of the counterflow. A direct effect is that the PEO itself retards the DNA motion, most likely due to a combination of hydrodynamic interactions and entanglement effects. The net result of these two opposing PEO-effects is that the center-of-mass velocity increases by a factor of about 2 upon addition of PEO. Circular plasmid DNA (13 kbp) can be electrophoretically trapped in gellan gels containing PEO.

Purpose: The purpose of this work is to investigate the various conditions affecting plasmid DNA separation in a particular polysaccharide gellan resin that forms hydrogels if a divalent ion is added. The gel can be reversed to solution by adding EDTA, which makes it a promising candidate for preparative electrophoretic separation of biomolecules.

Major Accomplishment: The electrophoretic migration of double-stranded T4 DNA (164 kilobasepairs) in gellan gels is shown to occur by changes in DNA coil conformation during the migration. Electroosmotic flow strongly reduces the electrophoretic stretching of the DNA, but has small effects on the path chosen by the DNA through the gel. The additive poly(ethylene oxide) (PEO) is shown to quench electroosmosis, leading to stronger stretching of the DNA, although PEO itself retards the DNA motion. The net result of these two PEO-effects is that the center-of-mass velocity increases by a factor of about 2 upon addition of PEO. It is shown that circular plasmid DNA (13 kbp) can be electrophoretically trapped in gellan gels containing PEO.

Impact: Electrophoretic separation of circular DNA is important because many of the vectors used for cloning and expressing DNA are circular. The electrophoretic separation of the different topological forms of circular DNA is determined by the conditions used, including: gel concentration, temperature, electric field strength, and buffer.

Future Plans: Future efforts will focus on using these techniques to process the DNA used in the Biotech Grain project, and in developing reference materials for DNA replication and processing.

Biocatalytic Systems

CSTL Program: Bio-Molecules and Materials
Authors: *M.P. Mayhew, Y. Tewari, M.J. Holden, D.T. Gallagher, V.L. Vilker, N. Smith (University of Maryland), and A.E. Roitberg (University of Florida)*

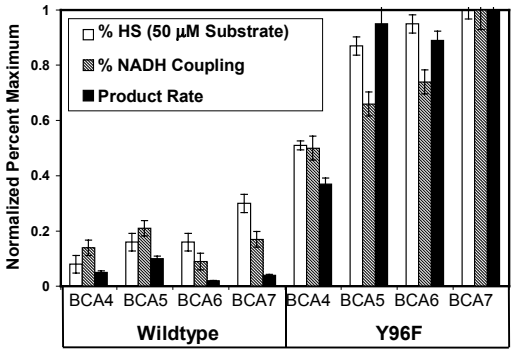
Abstract: Oxygenated benzocycloarenes (fused benzene-cycloalkane structures) are useful reactants in asymmetric syntheses for making pharmaceutical precursors. Toluene and naphthalene dioxygenases, and cytochrome P450 monooxygenases, either in the context of whole-cells or reconstituted enzyme systems, have been investigated as biocatalysts for hydroxylation of benzocycloarenes to regio- and stereospecific alcohols. With the advent of rapid methods for altering the properties of biocatalysts, industry is very interested in the development of screening tools that can be used to search for the best performance with respect to specific products. In this work, have shown that a simple spectrophotometric assay that takes advantage of the Soret band shift induced by substrate binding can predict the levels of P450 catalysis for each benzocycloarene-enzyme pair investigated.

Purpose: With the advent of rapid methods for altering the properties of biocatalysts, industry is very interested in the development of screening tools that can be used to search for the best performance with respect to specific products. It is the purpose of this project to use existing spectroscopy expertise and methods to develop screening methods that will accelerate commercial development of biocatalysts such as the mono- and dioxygenases.

Major Accomplishments: With the existing spectroscopy expertise at NIST, we have been developing screening methods that will accelerate commercial development of biocatalysts such as the mono- and dioxygenases. We have shown that a simple spectrophotometric assay that takes advantage of the Soret band shift induced by substrate binding can predict the levels of P450 catalysis for each benzocycloarene-enzyme pair investigated. This Soret shift assay is ready to be developed into a useful tool for identifying potential targets of P450 catalysis in a high-throughput screening format. It is appropriate for an industrial partner, with specific product targets, to pick up this development. We will continue to develop spectroscopic screening tools for rapid assessment of other biocatalyst properties such as cofactor utilization efficiency and kinetic limitations imposed by product inhibition.

Impact: Major chemical (Cognis, Dow) and pharmaceutical (Schering-Plough) companies have programs to develop asymmetric synthesis biocatalysts. This project shows that a simple spectroscopic method can be used to accelerate the screening of commercial biocatalysts such as the mono- and dioxygenases.

Future Plans: This Soret shift assay is ready to be developed into a useful tool for identifying potential targets of P450 catalysis in a high-throughput screening format. It is appropriate for an industrial partner, with specific product targets, to pick up this development. We will continue to develop spectroscopic screening tools for rapid assessment of other biocatalyst properties such as cofactor utilization efficiency and kinetic limitations imposed by product inhibition.


Normalized Percent Maximum
1
0.8
0.6
0.4
0.2
0
% HS (50 μM Substrate)
% NADH Coupling
Product Rate
BCA4
BCA5
BCA6
BCA7
BCA4
BCA5
BCA6
BCA7
Wildtype
Y96F

Biotech Grain Testing

CSTL Program: DNA Technologies
Authors: *M.J. Holden, V.L. Vilker, and J. Blasic (GeoCenters)*

Abstract: Grains grown in the US need to be characterized as to their content of genetic modification. NIST has been approached by the Grain Inspection, Packers and Stockyards Administration (GIPSA) of the USDA with a request to collaborate in establishing a US-based technical activity that can provide objective support for the GIPSA accreditation and performance verification programs. Private laboratories will be accredited by USDA to test US grains for the presence of biotech grains. We have worked with GIPSA and four US seed producers to gain access to their data bank of company proprietary genetic constructs that are used to make biotech crops. Material Transfer Agreements (MTA's) between NIST and the 4 companies (Monsanto, Aventis, Syngenta, and Pioneer) have been completed. We have participated in a preliminary round-robin screening study run by GIPSA to determine the lowest level of the Round-Up Ready trait that can be detected in corn by PCR methodology. Next year we will clone PCR products generated from trait materials made available through the MTA's into plasmid vectors for DNA sequencing. These vectors in turn will be available as positive controls for detecting some specific common traits in biotech grain. Next year we will also work to determine what a standard reference material should look like, based on input from USDA, plant genomic testing labs, seed producers, suppliers of reagents and other materials used in testing protocols.

Purpose: There is a critical need for DNA- and protein-based reference materials for validation of protocols and testing materials. The USDA has requested our help in developing methodology and reference materials to quantify the amount of genetically modified material in a sample of corn, and later cotton, wheat and soybeans. A detailed research program has been defined using input from a September 2000 meeting with our partners at GIPSA, the Institute for Reference Materials and Measurements (IRMM) of the European Commission's Joint Research Center.

Major Accomplishments: We have worked with GIPSA and four US seed producers to gain access to their data bank of company proprietary genetic constructs that are used to make biotech crops. Material Transfer Agreements (MTA's) between NIST and the 4 companies (Monsanto, Aventis, Syngenta, and Pioneer) have been completed. We have participated in a preliminary round-robin screening study run by GIPSA to determine the lowest level of the Round-Up Ready trait that can be detected in corn by PCR methodology.

Impact: Grains grown in the US need to be characterized as to their content of genetic modification. The Grain Inspection, Packers and Stockyards Administration (GIPSA) of the USDA have approached NIST with a request to collaborate in establishing a US-based technical activity that can provide objective support for the GIPSA accreditation and performance verification programs. The USDA will accredit private laboratories to test US grains for the presence of biotech grains.

Future Plans: Next year we will also work to determine what a standard reference material should look like, based on input from USDA, plant genomic testing labs, seed producers, suppliers of reagents and other materials used in testing protocols.

Quantitative PCR SRM

CSTL Program: DNA Technologie
Authors: *M.J. Holden, K. Cole, J.M. Butler, and J. Blasic (GeoCenters)*

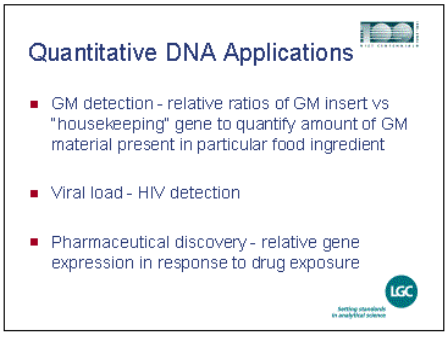
Abstract: The copying (amplification) of deoxyribonucleic acids (DNA) is used in many areas of biology and medicine to detect DNA sequences that are present in quantities too small to detect via any direct method. An enzymatic reaction, known as the Polymerase Chain Reaction (PCR) is the mechanism for amplification of the DNA. Laboratories that are detecting and quantifying the presence of DNA from genetically-modified crops in grain and food stuffs, estimating viral load in human disease process, detecting and quantifying pathogenic microbial DNA in environmental situations need a model DNA reference material in order to standardize equipment and protocols, much the same as the DNA reference materials are used in forensics and human identification. This model DNA system will be generic and can be used by any laboratory performing quantitative PCR and is not specific to detection of any particular gene.

Purpose: We will develop a model DNA reference material in order to standardize equipment and protocols used in detecting and quantifying the presence of DNA from genetically-modified crops in grain and food stuffs, estimating viral load in human disease process, detecting and quantifying pathogenic microbial DNA in environmental situations.

Major Accomplishment: Along with scientists from LGC, Teddington, UK, our staff participated in the planning and design of the first testing project for the new Biochemical Analysis Working Group of the CCQM – a model DNA system for quantitative PCR.

Impact: This reference DNA material will be unique in the world for the applications it is designed to serve.

Future Plans: A PCR protocol and DNA fluorescent probe need to be designed and tested in house using Real Time quantitative PCR testing methodology. An interlaboratory comparison will be conducted to evaluate the proposed SRM. We will verify of the correctness of the oligos by MALDI-TOF MS, and carry out self-assembly into full-length double-stranded sequences. This will be followed by insertion of the template sequence into plasmids, verification of template sequence order by DNA sequencing technology, plasmid production in large-scale bacterial culture, and purification of plasmids. Finally, quantification of plasmid preparations by various techniques and testing bulk plasmid production for integrity (gel electrophoresis) and sequence validity (DNA sequencing) will be done.



G-protein Coupled Receptor/Bacteriorhodopsin Chimeras: Grafting Segments from the Extracellular Surface of CCR5 onto the Transmembrane Helices of Bacteriorhodopsin Confers HIV-1 Coreceptor Activity

CSTL Program: Bio-Molecules and Materials

Authors: *K.D. Ridge; N.G. Abdulaev, T. Ngo, and R. Chen (CARB/UMBI); T.T. Strassmaier, and D.D. Oprian (Brandeis University, Waltham, Massachusetts); and H. Luecke (University of California, Irvine, California)*

Abstract: The chemokine receptor CCR5 is typical of the large family of G-protein coupled receptors (GPCR's) sharing the seven-transmembrane-helix structural motif that regulate a variety of sensory, hormonal, and neural responses. Currently, over 700 distinct sequences encoding putative GPCR's have been identified and many of these receptors play key roles in cardiovascular disease, diabetes, hypertension, AIDS, and a variety of sensory mental disorders. Components of the extracellular surface of CCR5 interact with certain macrophage-tropic strains of HIV-1 to mediate viral fusion and entry. In an attempt to mimic the viral interacting site(s) on CCR5, the amino-terminal and extracellular loop segments were grafted onto a seven transmembrane bacteriorhodopsin (bR) scaffold either singly, or in combination, to produce a set of CCR5/bR chimeras. The CCR5/bR chimeras were examined for stable expression, cellular localization, and/or gp160/120-mediated coreceptor activity. All of the CCR5/bR chimeras containing single or multiple replacements were stably expressed and many were transported to the cell surface. Chimeras containing the amino-terminal segment of CCR5, the third extracellular loop, the amino-terminus and third extracellular loop, or the entire extracellular surface, formed bR-like chromophores with all-*trans* retinal. Chimeras containing the amino-terminal segment of CCR5 exhibited significant chromophore-independent coreceptor activity while those chimeras possessing the entire CCR5 extracellular surface exhibited increased coreceptor function after retinal reconstitution. Like CCR5, chimera coreceptor function was dependent upon CD4, gp160/120 subtype, and tyrosine residues at positions 10 and 14.

Purpose: While chemical aspects of CCR5 have yielded to investigation, knowledge about the chemokine and HIV-1 binding sites is far from being understood at the molecular level. Further, the low natural abundance of CCR5 and difficulties in heterologous expression present a formidable challenge to attempts at high-resolution structure determination. The goal of this research is to develop alternative approaches to "express" functional GPCR domains in a form that is more amenable to study by X-ray crystallography or NMR.

Major Accomplishments: Biochemical studies show that some of the CCR5/bR chimeras effectively mimic CCR5 coreceptor function in forming a ternary complex with CD4 and HIV-1 gp160/120. Further, CCR5 segments that are primarily responsible for this interaction have been effectively delineated. These results indicate that segments from the extracellular surface of CCR5 can be functionally assembled on a "generic" seven transmembrane scaffold.

Impact: This approach may offer a promising strategy for appending functional domains from other GPCR's and expedite a structural understanding of the molecular basis for ligand binding and specificity among this class of receptors.

Future Plans: The immediate plans for this work are to overexpress some of the CCR5/bR chimeras in *H. salinarium* (the natural source of bR) and to crystallize the chimeras in the absence and presence of interacting proteins using the lipidic cubic phase approach.

IUPAC Report: Measurement and Analysis of Results Obtained on Biological Substances with Differential Scanning Calorimetry

CSTL Program: International Measurement Standards

Authors: *F.P. Schwarz; H.-J. Hinz (Institut für Physikalische Chemie, Westfalische Wilhelms-Universität, Schlossplatz 4/7 D-48149 Münster, Germany); A. Ginsburg, R. Szczepanowski, and M. Gonzalez (National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892-0342); S.A. Potekhin (Institute of Protein Research, Pushino, Moscow Region 142292, Russia); G. Rialdi, and F. Attanasio (Istituto di Studi Chimico Fisici di Macromolecole Sintetiche e Naturali IMAG-CNR 16149 Genova, Italy); and J.M.Sanchez-Ruiz, and B. Ibarra-Molero (Facultad de Ciencias, Departamento de Quimica Fisica, 18071-Granada, Spain)*

Abstract: Differential scanning calorimetry (DSC) has been widely used to determine the thermodynamics of phase transitions and conformational changes in biological systems including proteins, nucleic acid sequences, and lipid assemblies. This International Union of Pure and Applied Chemistry (IUPAC) report consists of recommendations for DSC measurement procedures, calibration procedures, procedures for testing the performance of the DSC instrument, analysis of the measurements, and the reporting of results. A test solution of (1-10) mg mL^{-1} lysozyme in HCl-glycine buffer at pH = 2.5 is recommended to check the operating performance of differential scanning calorimeters used for biological substances. Samples of the test solution were sent to six different DSC laboratories worldwide. Analysis of the data in terms of a two-state transition model yielded transition temperatures ranging from 329.4 K to 331.9 K and averaging 331.2 K and transition enthalpies ranging from 377 kJ mol^{-1} to 439 kJ mol^{-1} and averaging 405 kJ mol^{-1}. It is recommended that thermodynamic transition models only be applied to the analysis of the data if the model-derived transition temperatures and enthalpies are independent of the instrumental scan rate. Application of thermodynamic transition models involving two states, two states and dissociation, and three states to the heat capacity versus temperature data are described in detail

Purpose: One of the objectives of IUPAC is "to study topics of international importance to pure and applied chemistry that need standardization or codification". Accordingly, an IUPAC Biophysical Chemistry Commission sponsored project headed by F.P. Schwarz and H.-J. Hinz and consisting of a total working party of 10 scientists was formed to address the need for recommended procedures on the measurement and analysis of results obtained on biological substances with differential scanning calorimetry.

Major Accomplishments: The output of this project is the above IUPAC technical report published in Pure and Applied Chemistry and in The Journal of Chemical Thermodynamics and will be referenced in biological journals as well. A protein solution to check the operating performance of a differential scanning calorimeter was developed and tested successfully for use in six different international laboratories.

Impact: The test solution will be used to help validate the DSC measurement results and analysis performed in different laboratories. The report will help insure international uniformity in the nomenclature, analysis, and reporting of results obtained on biological substances with differential scanning calorimetry.

Future Plans: This activity is completed.

Single Molecule Fluorescence Detection of RNA Interactions

CSTL Program: Nanotechnology
Authors: *E.S. DeJong, and J.P. Marino; L. Locascio (839); M. Gaitan (812); A. Bardo, and L. Goldner (844)*

Abstract: We propose to provide the measurement infrastructure and modeling necessary to evaluate the structure and dynamics of single molecules *in-situ*. Our objectives are the development and demonstration of measurement tools for the manipulation, characterization and screening of single molecules. Specific goals include the development of 1) fluorescent single molecule (SM) detection and manipulation techniques to characterize assembly of multi-component RNA complexes and small molecule ligand binding to RNA; and 2) microfluidic technology to enable rapid, controlled mixture of solvents, buffers and macromolecules components at the nanoscale. We anticipate that the development of these measurement technologies will have dramatic impacts in the areas of bio- and nanotechnologies.

Purpose: Single molecule metrology lies at the frontier of measurement science and engineering and impacts the emerging fields bio- and nanotechnologies. Tools for the measurement and manipulation of single molecules have the potential to revolutionize the way medicine is practiced and materials are designed. New knowledge of genomic sequences and the structure and function of individual protein molecules make it increasingly important to have *general* measurement methods for nanoscale manipulation and ultrahigh-throughput screening of macromolecules, in order to accelerate development of biomaterials and pharmaceuticals.

Major Accomplishments: Fluorescent dye-labeled RNA molecules have been designed, synthesized and purified to perform single molecule fluorescence measurements with molecules immobilized on a surface. Initial measurements have been made (Figure 1) that demonstrate single molecule detection of RNA fluorescence emission on a surface. Using SM-FRET measurements, a standard macromolecular ruler is now being determined to allow estimation of distances in single molecules. Such measurements are being cross-validated with ensemble measurements on equivalent molecules.

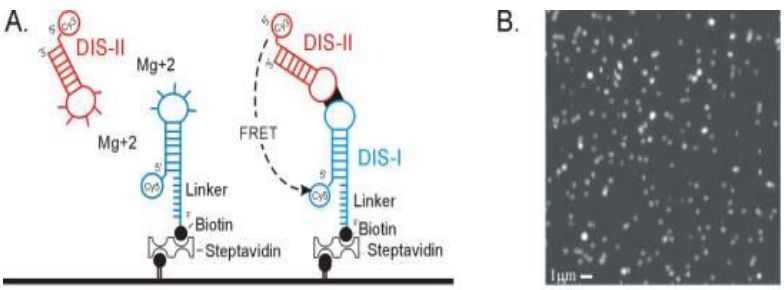


Figure 1: (A) Cartoon schematic for surface immobilization of RNA molecules for single molecule detection. Binding reactions of the immobilized RNA (Blue) will be measured using SM-FRET upon addition of the complementary RNA (Red). (B) Fluorescence emission intensity observed for individual RNA molecules.

Impact: SM measurements may enable new technologies in medicine, material science and molecular electronics (e.g. nanosensors, biomaterials, high-throughput screening, drug discovery and delivery). We aim to understand and harness the function of single macromolecules with the same ease that we currently understand and exploit the function of a single solid-state transistor.

Future Plans: We are in the first stage of this collaborative research effort involving the laboratories of CSTL, EEEL and Physics and anticipate ramping up over the next 5 years using support from the ATP and Competence programs. We hope the synergy that results from bringing individuals together with expertise in laser optics, microfluidics and biotechnology will allow us to make unique impacts in this emerging and exciting new field of metrology.

Structural and Biochemical Studies of Enzymes Along the Chorismate Pathway

CSTL Program: Bio-Molecules and Materials
Authors: *J.E. Ladner, and E. Eisenstein; J. Parsons, P.Y. Jensen, A. Pachikara, and K.E. Fisher (CARB/UMBI); and A. Howard (Illinois Institute of Technology)*

Abstract: Industrial production of drugs, new biopolymers and indigo dyes, is being facilitated through the use of products of the aromatic amino acid synthesis pathway of microorganisms.

The ability to manipulate this pathway through metabolic and protein engineering is saving non-renewable, petroleum-based feedstock chemicals. Our effort is focused on elucidating structure/function relationships of the enzymes involved in the chorismate metabolic pathway by solving their three-dimensional structures, modeling the mechanisms of the chemical transformations, and mapping pathway control nodes involved in the biocatalytic conversion of glucose to aromatic hydrocarbons. Only when the detailed three-dimensional structures are known for the enzymes can the precise enzymatic mechanisms and relationships between the structure and the physical properties be predicted, and modified. A state-of -the-art synchrotron X-ray source and cryo-crystallographic techniques were employed to obtain the highest possible resolution data and best three-dimensional structures. Specifically, the most recently elucidated structure was that of aminodeoxychorismate synthase, the product of the *pabB* gene. This structure provided the first look at the more complicated heterodimeric complex of the products of *pabA* and *pabB* which make up component I of *p*-aminobenzoic acid synthase, PABA synthase, in the bacterial metabolic pathway. Surprisingly, the structure revealed the presence of a bound tryptophan, linking the enzyme evolutionarily and structurally to the *TrpE* and*TrpG* complex of another branch of the chorismate pathway that leads to the synthesis of anthranilate. Crystallization of the *pabA-pabB* complex and the co-crystallization of a mutant of another enzyme in the pathway with its product in the active site are in progress.

Purpose: This work benefits pharmaceutical and chemical companies. The elucidation of structure/function relationships of the enzymes of these pathways enables more accurate mechanistic studies and rational design to improve the efficiency of these pathways to make these and similar chemicals.

Major Accomplishments: The work this year has yielded a better understanding of the chorismate pathway. We have solved the structure of aminodeoxychorismate synthase and have begun structural studies of a mutant of aminodeoxychorismate lyase. The structure of the synthase has also provided evidence of a stronger relationship between the various enzymes of the chorismate pathways than was previously thought to exist.

Impact: The synthesis of the aromatic amino acids is a very important area of study because these pathways do not exist in animals. They exist solely in bacteria, fungi, and higher plants.

Future plans: We plan to continue our studies of the enzymes along the chorismate pathway. Expanding our studies to incorporate active site ligands is, of course, a very obvious and important extension. This is now underway. We will also be looking at the complexes of the enzymes to understand more fully their precise roles and interactions.

Theoretical Analysis of Enzyme Structure, Reactivity, and Spectroscopy

CSTL Program: Bio-Molecules and Materials
Authors: *M. Krauss, and S. Worthington*

Abstract: Quantum calculation of stationary points on the reaction path of enzymes and the electronic structure and properties of enzyme active sites are pursued to explore the common electronic character of enzyme catalysis in metallic active sites and for enzymes in metabolic pathways. The methodology can be applied to the same or analogous enzymes within families of organisms or between different families to discern if there are common catalytic properties and discern functional properties of active site residues. The new methodology allows the quantum chemical calculation of hundreds of atoms and a realistic representation of the active site. A molecular dynamics/quantum protocol (MD/QM-EFP) has been developed to analyze enzyme reaction paths and is now being applied as widely as possible within the resources available. We are considering the following systems: chorismate mutase and 3-deoxy-D-arabino-heptulosonate-7-phosphate synthetase (DAHPS) in the shikimate pathway (and more generally organic chemistry in enzymes); glyoxalase I and other members of its functional super-family; bi-metallic enzymes, zinc lactamase, phosphotriesterase; and the repair enzyme, uracil-DNA-glycosylase.

Purpose: The aim of this research is to provide a simulation protocol to leverage one (or a few) experimental structures to determine the reaction path for native, mutant, or metal substituted enzyme systems and provide insight into the behavior of families of enzymes.

Major Accomplishments: Binding of substrate, product, and transition-state along the reaction path in chorismate mutase has been optimized for the native and some mutant enzymes using native and unusual substrates. The reaction path provides an understanding of the microscopic mechanism that identifies the role of each catalytic residue in the active site, beyond what is known from experimental results, and provides insight into the common catalytic features in non-homologous families of enzymes that perform the same chemistry in different organisms.

The product of the reaction catalyzed by uracil DNA glycosylase has been characterized by simulating the NMR shifts of product and active site residues in the product-enzyme complex. Spectroscopic properties (electronic, NMR, infra-red, etc) of molecules bound at various points along the reaction path can provide new insight into enzymatic behavior and provide the opportunity to correlate theoretical and experimental results.

Impact: The general methodology that has been developed in this research can be applied to any enzyme family. These tools allow more rapid and accurate simulations to be performed on many more enzymes of technological and medical interest through homologous structural modeling and reaction path optimization. This information is important in the pharmaceutical industries effort to more quickly develop useful products and drugs.

Future Plans: Collaboration with with B. Brooks and E. Billings (NHLBI/LBC, Computational Biophysics Section NIH) is planned to implement an effective ab initio quantum mechanics/molecular mechanics (QM/MM) approach to enzyme structure and reactivity. This effort, in combination with the effective fragment potential (EFP) approach developed at NIST, should significantly enhance the accuracy of modeling the conformation of enzyme active sites which is a crucial aspect of enzymic catalysis.

An Amino Acid Exchangeability Measure Derived from Experimental Data

CSTL Program: Chemical & Biochemical Data
Authors: *A. Stoltzfus; and L.Y. Yampolsky (CARB/UMBI)*

Purpose: The aim of this work is to derive useful information on the mean effects of amino acid exchanges from the results of experimental exchanges of amino acids under laboratory conditions. Measures of pairwise similarity between amino acids have many different applications in bioinformatics. The most commonly used measures, PAM and BLOSUM, summarize patterns of evolutionary divergence. When used to gauge the effects of amino acid change on proteins, however, evolution-based measures are problematic, because they reflect, not only effects of amino acid changes on proteins, but also the rates of underlying mutations necessary for evolutionary change to occur. Experimental data could be used to avoid mutational biases and thus clarify the mean effects of amino acid changes in proteins. The challenges are to find a sufficiently large set of data from experiments whose design does not impose strong biases, and to combine results from diverse studies.

Major Accomplishments: From 12 studies in which amino acid exchanges were engineered systematically, without imposing substantial biases, the results of 9334 single amino acid exchanges were tabulated. Results of diverse studies were transformed to a common scale using a statistical model relating frequency to severity of effect on activity, based on 7 studies that provided information on the activity of variant proteins relative to the wild-type. The resulting estimation of the mean effect EX_{ij} of changing amino acid i to amino acid j was evaluated by a jack-knife procedure in which a subset of the studies is used to predict the results of the remaining studies, with the predictive power being analyzed by logistic regression and compared to the power of other predictors. As shown in the table below, EX is a better predictor than all measures tested except for BLOSUM, even though most of these measures are based on far larger amounts of data from comparative sequence analysis (PAM, BLOSUM, Grantham) or from structural modeling (MJ).

	Combined surface & buried sites					Separate sites	
Predictor:	BLO	EX	PAM	MJ	1/Gra	EX	KG
Log Likelihood	-169.0	**-147.7**	-138.5	-91.4	-36.0	**-395.9**	-190.3

Separate EX matrices for surface and buried sites have been computed, and this pair of matrices outperforms a comparable pair of matrices based on evolutionary divergence (KG, Koshi-Goldstein). An asymmetric EX matrix, in which EX_{ij} and EX_{ji} are different, can also be computed: this matrix outperforms all other predictors tested, including BLOSUM.

Impact: This work is still unpublished, so there is no public impact as yet. The results have been presented in a seminar and have stimulated interest from a pharmaceutical company in the possibility of a CRADA to support further work.

Future Plans: EX will be incorporated into a general codon-based model for protein sequences that provides a formal basis for exploring the separate contributions of mutation and natural selection to protein evolution and to hereditary variation in human protein-coding genes.

Peptide Standard Reference Materials

CSTL Program: International Measurement Standards
Author: *E. Eisenstein*

Abstract: All products and processes in biotechnology rely heavily on accurate characterization and instrument calibration. Consequently, a tremendous investment of time and money is spent in industrial, government and academic laboratories on assessing and qualifying processes and products for biotechnology. These issues are especially important for protein and peptide biopharmaceuticals, for which relatively minor chemical modifications that might occur either in cells or upon storage could result in a complete loss of biological activity. In an effort to more accurately assess the physical and chemical properties of these materials, many laboratories are seeking standards to calibrate instrumentation for routine analysis of biological samples, as well as to benchmark the quality of novel biopharmaceuticals. A NIST-developed peptide SRM would aid these laboratories significantly by reducing the time and effort currently spent in assessing the competence, accuracy and reproducibility of preparing and characterizing proteinaceous biological substances. Three peptides have been designed to cover a useful range of size, weight, charge and solubility, prepared in small scale, and their physical and chemical properties characterized. The purity, characteristics, and stability of the three test samples suggest they would be useful as standards. Large-scale synthesis and purification are underway to provide this greatly needed standard for the biopharmaceutical industry.

Purpose: A NIST-certified peptide SRM would provide preparative and analytical laboratories in the public and private sector with key markers for calibrating a wide range of instruments, as well as for validating the physical and chemical properties of new biopharmaceuticals.

Major Accomplishments: Pilot studies on the optimal synthesis, purity, physical and chemical properties, and stability and shelf life have been performed on three designed model peptides. The high yields and purity of the samples indicate it is feasible to scale up synthesis to the gram level. Additionally, each peptides was studied in collaboration with a number of participating laboratories from the Association of Biomolecular Resource Facilities (ABRF) Quality and Assurance subcommittee. The peptides were characterized using ten bioanalytical approaches for which they will be utilized in research laboratories. The samples will be characterized for purity using reversed-phase HPLC, cation and anion exchange chromatography, and capillary electrophoresis. The extinction coefficients of the peptides will be determined using absorption spectrophotometry in combination with the results from amino acid analysis. Analytical molecular weights will be determined by matrix-assisted, laser-desorption, time-of-flight (MALDI-TOF) and electrospray-ionization (ESI) mass spectrometry. Finally, the samples will also be analyzed chemically by performing amino acid analysis, N- and C-terminal sequencing, and by protease digestion with Lys-C and trypsin, with fragment identification by mass spectrometry.

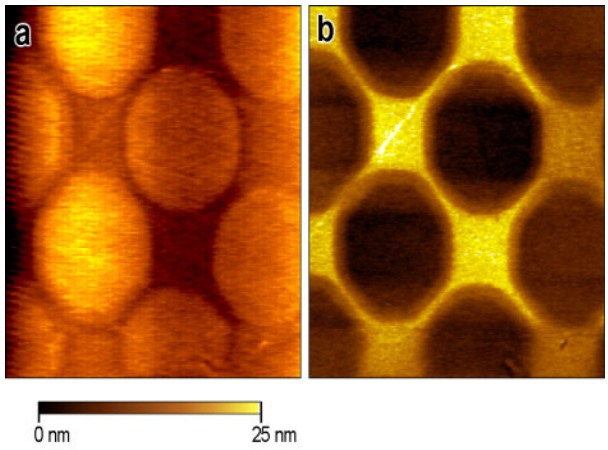
Impact: With a potential user base of up to 500 customers per year, validated peptide SRMs could save the biopharmaceutical industry alone about 1 M annually.

Future Plans: Large-scale synthesis of the three peptides is underway in an effort to produce 1-2 gram quantities for use as an SRM.

Molecular Recognition Force Microscopy

CSTL Program: Bio-Molecules and Materials
Authors: *J.T. Woodward, and J.T. Elliott*

Abstract: We are developing a molecular recognition force microscope (MRFM) capable of mapping the location of biomolecules on a surface with nanometer scale resolution at high speed. This is being accomplished through the modification of the electronics of a standard atomic force microscope (AFM) and by functionalizing the AFM tip with a ligand or antibody to the molecule of interest. We have modified the AFM electronics to operate in pulsed force mode which allows simultaneous topographic and adhesion force imaging. The figure shows the result of a test using nonspecific interactions between the AFM tip and hydrophobic (circles of alkanethiol) and hydrophilic (gold) areas. Figure (a) is a topographic image showing the microcontact printed alkanethiol circles and (b) is an adhesion image showing low adhesion on the alkanethiol and high adhesion on the gold, as expected.



Purpose: The ability to measure the density and distribution of biomolecules at the sub-micron length scale allows insights into cell-surface interactions that control cell adhesion and proliferation. This information is of critical importance to the developers of tissue-engineered medical products and cell-based devices.

Major Accomplishments: We have successfully integrated new electronics into our AFM system that allow the controlled oscillation of the AFM tip and the simultaneous measurement of topographic and adhesion force data. We have also developed techniques to use microcontact printing of proteins to make test samples for the MRFM.

Impact/Future: We are currently functionalizing AFM tips with antibodies to begin testing the ability to detect single molecule binding events.

New Optical Measurements for Characterizing Molecular Structure

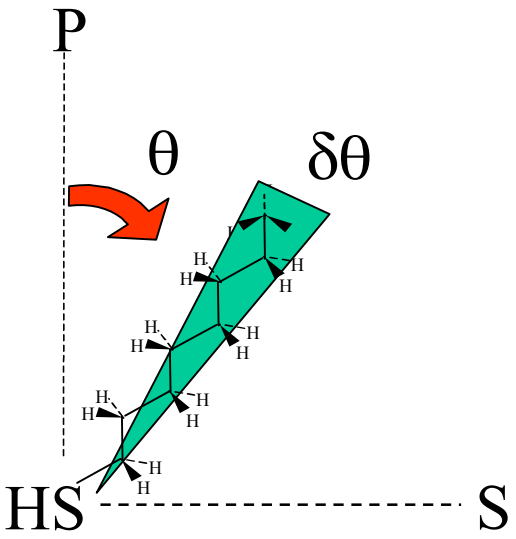
CSTL Program: Nanotechnology
Authors: *C.W. Meuse, and J.B. Hubbard; and R.D. Mountain (838)*

Abstract: The growth of biotechnological fields including pharmaceuticals, biosensors and disease diagnostics is limited by the ability to deduce structure-function information from ultrathin films and transmembrane proteins. For example, although the function of many surface immobilized proteins can be deduced using a variety of methods, there are no tools available to probe their structure. We are developing a set of infrared spectroscopy methods for characterizing the structure of molecules on surfaces or in the bulk. To date, we have experimentally obtained two independent polarized spectra from self-assembled alkanethiol monolayers on solid supports. We are developing a theory for using this information to obtain new structural parameters. These data are also being used to test the validity of molecular structural predictions. This approach may ultimately provide the structural information that is important for the rational development of biotechnological applications of ultrathin films and transmembrane proteins.

Purpose: We propose to broaden the use of infrared spectroscopy by developing user-friendly methods for structural analysis of molecules. These methods will be used to test techniques for predicting the structure and dynamics of macromolecules.

Accomplishments: We have developed new capabilities for measuring two different polarized spectra from self-assembled alkanethiol monolayers and methods for analyzing this data to deduce the statistical variance of the alkane chain molecular orientation (*see figure*). Our experimental results are in excellent agreement with predictions of molecular dynamics simulations of octadecanethiol monolayers.

Impact: Infrared spectroscopy is a powerful technique for assessing the presence of particular moieties in complex samples. However, it only extracts a small fraction of the structural information that is hidden because of spectral degeneracy. If IR can be expanded to estimate additional structural information, it will help validate molecular dynamics simulations. The combination of these two tools will aid in the screening pharmaceuticals, developing biosensors, analyzing tissue samples and diagnosing diseases.



Future Plans: We are developing new measurement techniques to provide the structural information necessary for the validation of structural predictions. These validated predictions will extend the limited available information and allow the rational development of ultrathin film and transmembrane protein applications biotechnology.

Protein Data Bank

CSTL Programs: Healthcare Measurements, Bio-Molecules and Materials
Authors: *T.N. Bhat, and G.L. Gilliland*

Abstract: The Protein Data Bank (PDB) is the single archive of biological macromolecular structures. The Research Collaboratory for Structural Bioinformatics (RCSB; http://pdb.nist.gov), has been fully responsible for its management since July 1, 1999. The RCSB has provided new integrated systems for collecting, distributing and querying the contents of the PDB. These systems were designed with the expectation that there would be a change in the quantity, quality, and content of new structures. It is the goal of the PDB to enable new science through effective use of structure data; however, the provision of the data may change. The biotechnology sector has generated vast amount of data and will continue to do so in the future. Consistent schema, uniform validation tools and standard database interfaces are needed to allow efficient queries and distribution of these data. Quality and uniformity are two major issues for dependable and reliable results. This work will focus on these issues.

Purpose: The Protein Data Bank (PDB) was established in 1971 with a handful of structures at Brookhaven National Laboratory and it is the single Internationally recognized archive and distribution center for the structural data. In 1998, the Research Collaboratory for Structural Bioinformatics (RCSB, http://pdb.nist.gov), a consortium consisting of Rutgers University, the San Diego Supercomputer Center (SDSC) at the University of California, San Diego (UCSD), and the National Institute of Standards and Technology (NIST) became the custodians of the PDB. The RCSB has provided new integrated systems for collecting, annotating, distributing and querying the contents of the PDB. The access and distribution of the PDB data is through the primary Web site at SDSC and through mirrors located at Rutgers University, NIST, and in other locations throughout the world. The PDB receives an average of 115,000 hits per day on the primary Web site alone. The PDB Web sites provide users with direct query and reporting capabilities using the underlying databases. The query capabilities are quite extensive, and have been substantially improved with the introduction of the Molecular Information Agent (MIA; http://mia.sdsc.edu/), which provides frequently updated links to a growing number of databases. Query across the complete PDB has nevertheless been limited by missing, erroneous, and inconsistently reported experimental data, nomenclature, and functional annotation. Inconsistency, in particular, reflects the evolution of experimental methods, functional knowledge of proteins, and methods used to process these data over the years. The result is that only PDB ID searches provide completely reliable results. The purpose of the data uniformity project that is underway is to address the non-uniformity in PDB data and thus enabling better query results for the customers.

Major Accomplishments: First stage of data uniformity work on approximately 15000 legacy Protein Data Bank (PDB) entries is completed and the results are now distributed to users (http://pdb.nist.gov). These data uniformity efforts focused on a broad range of issues like data quality, data standards and data exchange. To further enhance the interoperability of the structural data, CORBA standard is proposed and established (OMG Document Number: lifesci/2000-11-01) by the PDB.

Impact: The data uniformity effort ensures quality and standards to permit reliable data query and data exchange for millions of PDB customers. Our publication that described the Protein

Data Bank effort (Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. , 2000), The Protein Data Bank. *Nucleic Acids Res.*, **28,** 235-242) has been rated by the Science watch as the second highest impacting publication of the year 2000.

Future plans: Our efforts on data uniformity and data standards is an on going long range program. The total number of entries held by PDB is expected to double in the next three years and this increase is expected to further underscore the importance of data uniformity.

Bioinformatics Software Resource

CSTL Programs: Healthcare Measurements, Bio-Molecules and Materials
Authors: *T.N. Bhat, and G.L. Gilliland*

Abstract: Data standards and software tools are the two key elements of information technology. In the area of Bioinformatics and Macromolecular structures extensive data uniformity efforts are underway (Bhat et. al, Nucleic Acid Research, 2001, Vol. 29, No. 1). Further a modern resource for data archival and distribution have also been established (http://www.rcsb.org/). However, there is an urgent need for similar collective effort from the point of view of software and tools for Bioinformatics and for macromolecular structural research. The approach taken to provide this resource involves a) the development of a website that utilizes a database of web links and library of software and benchmark data for Bioinformatics/Computational Biology with particular emphasis on database development and data mining, b) development of standard application program interface (API) for many of the commonly used tasks in Bioinformatics. Format conversions, standard object definitions and CORBA standards are some of the tasks of interest for API development.

Purpose: Data standards and software tools are two critical elements of information technology. Many of the efforts (http://pdb.nist.gov) related to interoperability in Bioinformatics are focused on data. However, a successful data exchange also requires a resource for standard software tools for parsing and analyzing the information. For this purpose a Bioinformatics Software Resource (BISR, http://bioinfo.nist.gov/BISR) has been developed. At present BISR provides key-word searchable web links for sites that distribute Bioinformatics software. In the future BISR also plans to establish a database with features to archive and distribute software. Through this effort, our goal is to catalyze and contribute to the development of standard application program interface (API) for format conversions and data exchange.

Major Accomplishments: A prototype for a web resource for software in Bioinformatics is developed (http://bioinfo.nist.gov/BISR/). Long range objective of this web resource is to provide a stable archival and distribution facility for publicly available software in Bioinformatics. As a first step, now our web resource provides key-word searchable links to web resources that distribute software.

Impact: Up until now most of the efforts in uniformity and standards in structural Biology have focused on data. Recent developments in technology have created an urgent need for similar efforts in software standards as well. The proposed resource is expected to support and catalyze such activities.

Future Plans: We plan to enhance this resource to allow archival and distribution of public domain software and to develop application program interface to foster interoperability among software.

AIDS Related Structural Database

CSTL Programs: Healthcare Measurements, Bio-Molecules and Materials
Authors: *T.N. Bhat, G.L. Gilliland, and V. Ravichandran*

Abstract: AIDS is a growing epidemic and several industries are actively working on developing structure-based drugs to control the disease. However, structural information on proteins that are potential targets for AIDS are scattered and data exchange between these resources is difficult. A reliable archival and annotated data distribution facility for the structural data on AIDS related proteins are critical for a sustained and successful drug development activity. For this purpose a specialized resource is being developed in collaboration with National Cancer Institute.

Purpose: Structure based drug design against AIDS is one of the most successful AIDS related research. However, due to viral sequence mutations, in spite of the availability of several potent structure based anti-viral drugs, AIDS is a growing epidemic. For this reason sustained research and drug development is needed to combat AIDS. To support such an effort, there is an urgent need for a primary archival and distribution resource for structural information on AIDS related proteins. In collaboration with National Cancer Institute, a structural database for AIDS related proteins is planned http://srdata.NIST.GOV/hivdb/hivdb2.ASP. This structural database will receive, annotate, archive and distribute all structural data for proteins related to AIDS. HIV protease, integrase and RT are few such proteins to be included in this database. During the annotation of the data, special efforts will be made to focus on information relevant to drug resistance.

Impact: Availability of a specialized resource for a very important Healthcare topic like AIDS will permit faster and better reliable access for data related to drug design and development. It will also provide improved resources for analyzing drug resistance for drugs that are currently used to treat AIDS.

Future Plans: We plan to further develop this resource from the point of view of data uniformity across all aspects of HIV related proteins, with particular emphasis to drug design. This plan includes developing tools to navigate through the information available on drug enzyme interactions. One of our staff is working on developing a histogram matching technique for navigating through the different aspect of drug enzyme interactions.